

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-116094  
 (43)Date of publication of application : 06.05.1998

(51)Int.Cl. G10L 3/00  
 G10L 3/00  
 (21)Application number : 09-265959 (71)Applicant : LUCENT TECHNOLOG INC  
 (22)Date of filing : 30.09.1997 (72)Inventor : UU CHOU  
 JUANG BIING HWANG  
 KAWAHARA TATSUYA  
 CHIN-FUI LEE

(30)Priority  
 Priority number : 96 724413 Priority date : 01.10.1996 Priority country : US  
 96 771732 20.12.1996 US

## (54) METHOD AND DEVICE FOR VOICE RECOGNITION

## (57)Abstract:

PROBLEM TO BE SOLVED: To efficiently conduct conversation voice recognition with a high degree of freedom while accepting freer conversation than the conversation based on a fixed and formalized grammar. SOLUTION: A key phrase detector 11 detects plural key phrases based on the set of the phrase part grammar of the intrinsic condition of the conversation. Then, a key phrase verifier 12 assigns reliability scales to these key phrases, the scales are compared with threshold values to verify the key phrases and a set of verified key phrase candidates is obtained. Then, a sentence hypothesis generator 13 couples the candidates based on a prescribed (task intrinsic, for example) meaning information 25 and a sentence hypothesis 14 is obtained. Lastly, a sentence hypothesis verifier 14 verifies the sentence hypothesis, verified sentence hypothesis is generated and a final recognition result is obtained. In the conversation base system, individual phrase part grammar is used in accordance with the specific state of the conversation in a large scale task (i.e., against 'subtasks').

## LEGAL STATUS

[Date of request for examination] 05.04.1999

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

(10)日本国特許庁 (JP) (12)公開特許公報 (A)

(11)特許出願公開番号

特開平10-116094

(43)公開日 平成10年(1998)5月6日

(51)Int.Cl.\* 識別記号 FI  
 G10L 3/00 G10L 3/00 S61G S61C  
 561 531

審査請求 未請求 請求項の数32 OL (全 12 頁)

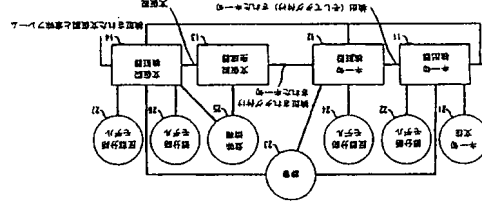
(21)出願番号 特願9-265959  
 (22)出願日 平成9年(1997)9月30日  
 (31)優先権主張番号 08/724413  
 (32)優先日 1996年10月1日  
 (33)優先権主張国 米国 (US)  
 (31)優先権主張番号 08/771732  
 (32)優先日 1996年12月20日  
 (33)優先権主張国 米国 (US)  
 (71)出願人 59607259  
 ルーセント テクノロジーズ インコーポ  
 レイテッド  
 Lucent Technologies  
 Inc.  
 アメリカ合衆国 07974 ニュージャージー  
 ー、マレーヒル、マウンテン アベニュー  
 600-700  
 (72)発明者 ウー チョウ  
 アメリカ合衆国、07922 ニュージャージー  
 ー、パークレーン ハイツイ、グリーンブライ  
 アー ドライブ 22  
 (74)代理人 弁護士 三橋 弘文  
 最終頁に続く

## (54) 発明の名称 音声認識方法および音声認識装置

## (57) 【要約】

【課題】 固定した形式的な文法に基づいて受容されるよりも多くの発話を受容される、効率および自由度の高い会話音声認識を実現する。

【解決手段】 キー句抽出器 11で、会話の状態に固有の句部分文法の集合に基づいて複数のキー句を抽出する。次に、キー句検証器 12で、これらのキー句に属する性質尺度を割り当て、その信頼性尺度をしきい値と比較することによってそれらのキー句を検証し、検証済みキー句候補の集合を得る。次に、文脈生成器 13で、検証済みキー句候補を、所定の (例えばタスク固有の) 意味情報 25に基づいて結合して文脈説を得る。最後に、文脈検証器 14で、これらの文脈説を検証して、検証済み文脈説を生成し、最終認識結果を得る。会話ベースのシステムでは、大規模なタスク内の会話の特定の状態に応じて (すなわち「サブタスク」に対して)、個別の句部分文法を使用することが可能である。







12によって検証され、信頼性尺度が割り当てられる。上記のように、このプロセスは、これがなければ起こり得る多くのフォールスアラームを除去する。実施例のキ一句検出器12は、当業者に周知の「反部分語モデル」を用いて、認識されたキ一句の各部分語をテストする。部分語レベルの検証の組合せからなる。キ一句検出器12は辞書23、部分語モデル22および反部分語モデル24を使用する。これらは、例えば、最小検出誤差(MVE (minimum verification error))基準を用いてトレーニングされたものである。

【0021】図1の実施例の第3の構成要素は文脈生成器13である。これは、例えばタスク固有の意味情報25を用いて、検証されたキ一句候補を、1つ以上の文脈へと結合する。例えば、I. Kawahara et al., "Concept-Based Phrase Spotting Approach for Spontaneous Speech Understanding", Proc. IEEE-ICASSP, pp. 291-294, 1996, に記載されたようなスタック復号器を用いて、意味制約を満たす最適な仮説を探索することができ

る。

【0022】最後に、文脈検出器14によって、音響的かつ意味的に最良の意味仮説が検証され、最終出力(すなわち、少なくとも1つの検証された文脈)が生成される。文脈検出器14は、意味情報25、辞書23、部分語モデル26および反部分語モデル27を使用する。キ一句に付けられた意味タグが、キ一句検出器1によって提供された意味仮説生成器13によって使用されるため、検証された文脈は本質的に、直接に対応する「意味」を有し、それにより、個々のアプリケーションによる必要に応じた意味フレーズの生成が可能となる。

【0023】キ一句検出器11は、キ一句検出を実行する。これは、会話状態に依存する特定のサブタスクに基づくことが可能である。具体的には、各サブタスクごとに、キ一句パターンが1つ以上の決定性有限状態文法として記述される。これは、実施例では、キ一句検出器11によってキ一句文法21から選択される。これらの文法は、タスク仕様から直接に入手により導出することも可能であり、あるいは、当業者に周知の従来の学習手続きを用いて、小さいコーパスから自動的にまたは半自動的に(すなわち、人間の支援のもとで)生成することも可能である。

【0024】一般に、キ一句は、従来のキ一句ワークロードに比べて、"at the"や"near"のような機能語を含む。これにより、従来のキ一句ワークロードのマッチングに比べて、より安定なマッチングが可能となり、検出精度が改善される。(例えば、前掲のI. Kawahara et al., "Concept-Based Phrase Spotting Approach for Spontaneous Speech Understanding"を参照。)いずれのキ一句にも含まれないがしばしばキ一句に伴う充填句も定義され、埋め込まれたキ一句を含む句パターンを形成するために使用

される。

【0025】特に、キ一句および充填句の文法はネットワークへとコンパイルされる。このネットワークにおいて、キ一句は繰り返り現れ、ガーベジモデルがキ一句の出現の間に埋め込まれる。しかし、注意すべき点であるが、単純な繰り返りは曖昧さを生じる可能性がある。例えば、目の繰り返り検出が許容される場合、"twenty four"と"twenty"を区別することはできない。従って、不可能なキ一句の結合を禁止する追加の制約も組み込む必要がある。

【0026】従って、図2に、許容される結合および反復を有するキ一句部分文法オートマトンのネットワークからなる。このようなオートマトンは、結合重みを評価することによって、確率的言語モデルへと容易に拡張することができる。このようなモデルを使用することにより、文レベルの文法と比べてあまり複雑にならずに、適用範囲が広がる。

【0027】例として、図2に、単純化した(すなわち、簡略化した)句ネットワークの例を示す。これは、「データ取得」サブタスクに適用された場合に、図1の実施例のシステムのキ一句検出器11によって使用されることが可能である。このネットワーク例の完全な実現により、曜日、月、日、および年の要素的に仕様の反復が、適当な制約のもとに許容される。(このように完全な実現の全探索は99語である。)この特定のサブタスクでは、キ一句は組み込まれない。

【0028】さらに具体的には、ここに記載する本発明の実施例によって採用されている検出方法は、フォワードバックワード2パス探索に基づくものである。これは、例えば、W. Chou et al., "An Algorithm of High Resolution and Efficient Multiple String Hypothesis Resolution for Continuous Speech Recognition Using Inter-Word Models", Proc. IEEE-ICASSP, volume 2, pp. 153-156, 1994, に記載されている。本発明の別の実施例では、代わりに、当業者に周知の1パス検出法を使用することも可能である。

【0029】A. 設定スタック復号器 (例えば、前掲のI. Kawahara et al., "Concept-Based Phrase Spotting Approach for Spontaneous Speech Understanding"に記載されているもの) は、N番目までの最良ストリングとして得られるN個の最良仮説は一般に、1〜2語が過ぎ検わった類似の単語列である。本発明の目標は、(入力候補全体に基づいてストリング仮説を生成することではなく)入力候補の一部に基づいてキ一句候補を識別することであるので、仮説を延長しても既に延長された仮説と同じ仮説になる場合にはその仮説は捨てられる。

【0030】特に、本実施例のスタック復号器は、キ一句ネットワークのマーキング(merging)状態にマークを付けることによって実現される。当業者には周知のよう

に、マーキング状態は、キ一句あるいは充填句が終了し、さらに延長する次の(すなわち新たな)句の最初に導入することになるノードに対応する。

【0031】スタック復号器によって「ポップ」される仮説に、出力されるべき完全な句であるというタグが付いている場合、本発明の手続きは、もう1語だけの句を延長し、その句を最良延長と並べる。このノードに、以前のいずれかの仮説が同じ時点で到達している場合、検出した句を出力した後現在仮説は捨てられる。そうでなければ、その時点は、その後の探索のためにマークされる。

【0032】注意すべき点であるが、この検出手続きは、冗長な仮説延長のない効率的なものであり、スコア順に、正しいN番目までの最良のキ一句候補を生成する。本発明のさまざまな実施例によれば、この手続きは、所望の個数の句を生成したことに基いて、あるいは、あるスコアしきい値に基いて、終了することも可能である。例えば、仮説のスコアが、最良スコア仮説の0.99倍より小さい値に到達したときに、検出を終了することも可能である。

【0033】キ一句検証と信頼性尺度 図1の実施例のシステムのキ一句検出器12は、部分語レベルのテストに基づいて、検出された句の検証を行う。具体的に、与えられた句の各部分語nに対して、検証スコアは、次式のような従来の尤度比(LR (likelihood ratio))テストに基づいて計算される。

$$\log LR_n = (\log P(O|\lambda_n^*) - \log P(O|\lambda_n)) / \log P(O|\lambda_n^*) \quad (2)$$

注意すべき点であるが、式(2)の第1項は認識スコアそのものである。上記の計算の効果は単に、計算されるスコアを反部分語モデルのスコアだけずらし、その結果を正規化することである。

【0036】キ一句検出器12は、検出された各キ一句ごとに、対応する部分語レベルの検証スコアを組み合わせ

$$CM = \left( \log LR_1, \dots, \log LR_n \right) \quad (3)$$

信頼性尺度(CM)が、ある所定のしきい値を超える場合に、与えられたキ一句は承認される。実施例では、しきい値の値は、例えば-0.15に設定される。

【0037】本発明のさまざまな実施例において、さまざまな信頼性尺度関数を使用することができ、例えば、第1の明示的な信頼性尺度CMは、フレーム継続

$$CM_1 = \frac{1}{L} \sum_{n=1}^L (l_n \cdot \log LR_n) \quad (4)$$

上記の式で、 $l_n$ は、部分語nの継続時間を表し、Lは句の全継続時間である。すなわち、 $L = \sum l_n$ である。

【0038】第2の明示的な信頼性尺度CMは、部分語セグメントによる正規化に基づく。特に、これは、与えられたキ一句のすべての部分語の対数尤度比の単なる平均である。(一実施例では、句セグメンテーション後

\*  $LR_n = P(O|\lambda_n^*) / P(O|\lambda_n)$  (1)

ただし、Oは、観測フレーズの列を表し、 $\lambda_n^*$ および $\lambda_n$ は、それぞれ、部分語nに対する正しい部分語モデルおよび反部分語モデルを表す。(部分語モデルは部分語モデル22から得られ、対応する反部分語モデルは反部分語モデル24から得られる。)認識の結果として、観測Oは、部分語nに対して、ビタビアルゴリズムにより並べられ、最良P ( $O|\lambda_n^*$ ) およびP ( $O|\lambda_n$ ) が得られる。(ビタビアルゴリズムは、当業者に周知の従来のスコアリング方法である。)

【0034】各部分語モデルに対して、対応する反部分語モデルは、逆同しやすしい部分語クラスをまとめること(クラス化)によって構成される。各反部分語モデルは、対応する部分語モデルと同じ構造、すなわち、同じ個数の状態およびミクスチャを有する。反部分語モデルは、特定の部分語の検証専用であるため、反部分語モデルをリファレンスとして使用して復号を行うことにより、部分語モデルの無制約な復号を行うのに比べて、弁別性が改善される。こうして、システムは、認識器によってなされる置換誤りを拒絶する能力が増大する。この(検証)ステップでは、文脈独立な反部分語モデルを使用することも可能であるが、認識ステップは、文脈依存の部分語モデルを用いて実行される。

【0035】特に、上記の式(1)の対数をとり、その結果を、観測Oの継続時間長 $l_n$ に基づいて正規化することにより、量 $\log LR_n$ が次のように定義される。

$$\log LR_n = (\log P(O|\lambda_n^*) - \log P(O|\lambda_n)) / \log P(O|\lambda_n^*) \quad (2)$$

※させることによって、信頼性尺度(CM (confidence measure))を計算する。例えば、検出されたキ一句がN個の部分語を含む場合、このキ一句に対する信頼性尺度は、対応するN個の尤度比の関数とすることが可能である。具体的には、次のようになる。

$$CM = \left( \log LR_1, \dots, \log LR_n \right) \quad (3)$$

★時間による正規化に基づく。特に、これは、正しい部分語モデルに対して得られるビタビスコアと、対応する反部分語モデルに対して得られるビタビスコアの差に等しい。すなわち、次のようになる。

$$CM_1 = \frac{1}{L} \sum_{n=1}^L (l_n \cdot \log LR_n) \quad (4)$$

に単語間文脈情報が失われるため、最後の部分語に対して特別の考慮がなされる。)すなわち、次のようになる。

$$CM_2 = \frac{1}{L} \sum_{n=1}^L (l_n \cdot \log LR_n) \quad (5)$$

【数2】

【数1】

【数1】

$$CM_1 = \frac{1}{L} \sum_{n=1}^L (l_n \cdot \log LR_n) \quad (4)$$

【数2】

【数1】

\*する場合には、この句の前半（その確証スコアは低くなる可能性が高い）に注目するのが有効である。

$$CM_2 = \frac{1}{N} \sum_{n=1}^N \log LR_n \quad (5)$$

【0040】このように、低い信頼性レベルの部分語に注目するために、各部分語ごとに正規分布を仮定することによって、対数尺度を調整することが可能である。具体的には、部分語HMMのトレーニングで用いたサンプルを使用して、各部分語ごとに対数尺度比の平均および分散を計算する。その後、対数尺度比が、期待される平均より小さい部分語のみを含む和を実行することによって、CMsを計算することができる。すなわち、次のようになる。

【数3】

$$CM_3 = \frac{1}{N} \sum_{n=1}^N \left\{ \log LR_n, \log LR_n < 0 \right\} \quad (6)$$

その他の場合

※イド関数を用いる。この例示的な信頼性尺度は、最小誤り率基準でトレーニングするための損失関数として用いられる。すなわち、次のようになる。

【数4】

$$CM_4 = \frac{1}{N} \sum_{n=1}^N \frac{1}{1 + \exp(-\alpha \cdot \log LR_n)} \quad (7)$$

発生装置13はキー句候補器12によって生成された候補語のみキー句候補を意味情報25に基づいて1個以上の文脈と組み合わせる文脈解析を実行する。一実施例では、句候補のLR (left-to-right) トレリスを使用することが可能な1次元RL (right-to-left) 探索を用いられる。別の実施例では、効率的探索アルゴリズムを用いることも可能である。トレリス解析は計算量が多いため、さらに別の実施例ではラティス解析法を採用する。これは、トレリス解析よりわずかに精度が低くなるだけである。ラティス解析法は、音響スコアと、提供される意味制約情報（キー句タグの許容される組み合わせを指定する）に基づいて、句候補を結合する。キー句抽出のためのフォワードバックワード探索によって与えられるスコアを音響スコアとして用いることが可能である。

【0044】最も可能性の高い文脈を効率的に見つけるためには、スタック復号探索法を採用すると有効である。この方法は、連の部分仮説を反復的に生成し、完全な文脈仮説が生成されるまで、各反復において最良の利用可能な部分仮説を延長する。

【0045】具体的には、現在の「最良の」部分仮説を (w1, w2) とし、新たな仮説が句w3を連結することによって生成されると仮定する。新たな仮説 (w1, w2, w3) に対する評価関数は、完全な入力発話h0に對する上開スコアからのずれ（オフセット）として以下のように計算される。

【数5】

【0043】【文解折】図1の実施例のシステムの文脈

$$\begin{aligned} \hat{f}(w_1, w_2, w_3) &= h_0 - (h_0 - \hat{f}(w_1)) - (h_0 - \hat{f}(w_2)) - (h_0 - \hat{f}(w_3)) \\ &= \hat{f}(w_1, w_2) - (h_0 - \hat{f}(w_3)) \end{aligned} \quad (9)$$

ただし、 $\hat{f}(w_i)$  は、抽出された句w<sub>i</sub>に対する評価の結果である。初期仮説は  $\hat{f}(\text{null}) = h_0$  である。新たな句が追加されるごとに、オフセットが計算される。上開h0は、認識プロセスのフォワードパスで計算される。

【0046】上記の方法は、例えば、W. A. Woods, "Optimal Search Strategies for Speech Understanding Control", Artificial Intelligence, 18:295-326, 1982, に記載されているような不足法 (short-fall method) に基づいている。注意すべき点であるが、この評価はA・探索である。しかし、探索を効率的に導くこの方法の発見能力はやや限定されたものとなる可能性がある。抽出ベースの解析段階では特に、入力発話全体が扱われることを仮定しないため、敬語の短い仮説が誤って受容される可能性が高い。従って、発話でスキップされた部分を評価することが有効となる。そのため、具体的に、本発明の一実施例によれば、スキップ量に比例する一様な罰金値をオフセットとして追加することが可能である。もちろん、この近似は粗雑であるため、改善の探索となる可能性がある。従って、これを補うために、でべん句（無音を含む）を生成することが好ましい。（一実施例では、これらの仮説は、キー句除根プロセス中に生成することも可能である。）

【0047】【文解折】図1の実施例のシステムの文脈除根器14は、認識出力の最終判定を行う。実施例では、大体的音響情報および大体的意味情報の両方を使用し、それぞれ入力発話全体に適用される。キー句除根プロセスは局所的な判定のみを用いたが、文脈除根プロセスはこれらの局所的な結果を組み合わせて、従来の発話除根と同様の効果を実現する。しかし、抽出ベースの認識プロセスは、多数の予測しないキャリア句を含む場合でも入力発話を受容することが多いことに注意すべきである。

【0048】具体的には、文脈除根器14によって実行される音響除根プロセスは、与えられた文脈仮説が十分に一致することを保証するために、入力発話全体の両スコアリングを行う。この再スコアリングは、部分語モデル26、反部分語モデル27、および辞書23を用いて行われる。この段階で適用される部分語モデル（すなわち、部分語モデル26）の集合および対応する反部分語モデル（すなわち、反部分語モデル27）の集合は、キー句抽出器11およびキー句除根器12によって使用されるもの（すなわち、部分語モデル22および反部分語モデル24）よりも精度が高い。こうして、より高い精度の音響再スコアリングが実行される。

【0049】一方、意味除根プロセスは、与えられた各

文脈の意味的「完全性」を評価する。例えば、本発明の一実施例によれば、意味除根は、ある構成要素が意味的に「合法」かどうかのみを指定する単純な意味制約情報に基づいて実行される。このような場合、文脈除根器14の意味解析部分は、例え、与えられた文脈仮説の意味表現が完全であるかどうかを判断する。しかし、注意すべき点であるが、会話ベースのアプリケーションでは、例えば、不完全な発話にしばしば遭遇する。例えば、ユーザはただ「August」（8月）と書くだけで、その月の特定の日を指定しないことがある。一般に、こうした「不完全な」発話も同様に受容すべきである。

【0050】従って、本発明の一実施例によれば、文脈除根器14は、与えられた文脈仮説の意味表現を完成しておらず、かつ、ほとんどの入力セグメントが欠落比デストで拒絶された場合にのみ、その文脈仮説を拒絶する。この組合せ「デスト」は、例えば、満足な文脈仮説に遭遇するまで、各文脈仮説に適用することが可能である。

【0051】しかし、本発明の別の実施例では、さらに一般的な確率的意味モデルを、文脈除根器14で用いることが可能である。このような場合、各文脈仮説について、音響スコアとともに意味スコアを求め、組み合わせたスコアを用いて、最終認識結果として出力すべき候補とされた文脈仮説を選択することが可能である。

【0052】本発明のさらに別の実施例では、意味的候補のみまたは音響的候補のみ（両方ではない）を、文脈除根器14で実行することが可能である。例えば、さらに高い精度の部分語および反部分語のモデルが利用可能な場合には、入力発話の音響再スコアリングを実行することがあり効果がない。従って、この場合、意味除根のみを実行して、単に、与えられた文脈仮説が意味表現を完成していることを検証するか、あるいは、最早の意味モデルを用いている場合には、検証済み文脈仮説が最終検証結果として判断されるものになる意味スコアを生成する。

【0053】【配記】説明を明確にするため、ここに記載した本発明の実施例は、個別の機能ブロックからなるものとして表した。これらのブロックによって表される機能は、共用あるいは専用のハードウェアの使用によって提供することが可能である。ハードウェアには、ソフトウェアを実行することが可能なハードウェアが含まれるが、これに限定されるものではない。例えば、ここに記載した構成要素の機能は、単一の共用プロセッサによって、あるいは、複数のプロセッサによって提供することが可能である。本発明の実施例は、デジタル信号プロセッサ (DSP) ハードウェア、上記の動作を実行するソフトウェアを格納する読み出し専用メモリ (ROM)、および、結果を格納するランダムアクセスメモリ

(RAM) からなることが可能である。超大規模集積 (VLSI) ハードウェアや、カスタムVLSI回路を汎用プロセッサやDSP回路と組み合わせたものも可能である。

【0054】また、「キー句検出器」、「キー句検証器」、「文仮脱生成器」、および「文仮脱検証器」という用語は、対応する機能を実行する任意のメカニズムを含む。

0055]

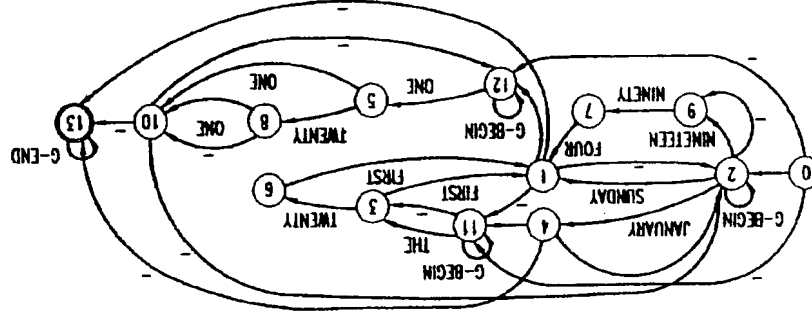
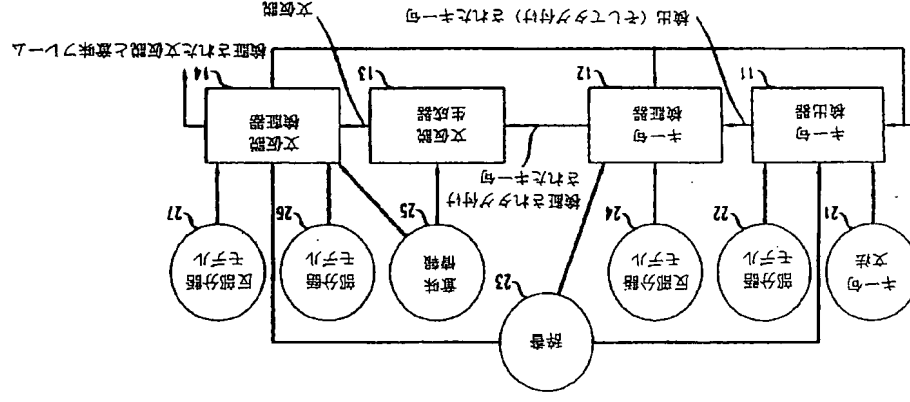
【発明の効果】 以上述べたごとく、本発明によれば、固  
定した形式的文法に基づいて受容されるよりも多くの  
発話が受容される、効率および自由度の高い会話音声認  
識が実現される。

「図面の簡明な説明」

【図 1】本発明の実施例による音声認識および音声発話の理解を実行するシステムの図である。

【图2】

【圖 1】



フロントページの続き

(71)出願人 596077259

600 Mountain Avenue,  
Murray Hill, New Je  
rsey 07974-0636 U. S. A.

(72)発明者 ビン・ホワン ジャン  
アメリカ合衆国、07059 ニュージャーシー、ウォレン、サウス レーン 8

(72) 発明者	かわはら たつや	(12)	特開平10-116094
	京都府京都市伏見区東奉行伏見御堂122		
(72) 発明者	チンーフイ リー		
	アメリカ合衆国、07574 ニュージャージー		
	ー、ニュー プロビデンス、ラニーメデ		
	パークウェイ 118		

# **NETWORK INTERACTIVE USER INTERFACE USING SPEECH RECOGNITION AND NATURAL LANGUAGE PROCESSING**

Publication number: JP2002524806T

Publication date: 2002-08-06

Inventor:

Applicant:

Classification:

- International: G06F3/16; G06F13/00; G06F17/27; G06F17/28; G10L15/18; G10L15/26; H04L29/06; G06F3/16; G06F13/00; G06F17/27; G06F17/28; G10L15/00; H04L29/06; (IPC1-7): G06F17/28; G06F3/16; G06F13/00

- European: G06F3/16; G06F17/27; G10L15/18C2; G10L15/26A; H04L29/06

Application number: JP20000569391T 19990908

Priority number(s): US19980150459 19980909; US19980166198 19981005; WO1999US20447 19990908

Also published as:

WO0014728 (A1)  
EP1110206 (A1)  
US6434524 (B1)  
EP1110206 (A0)  
CA2343150 (A1)

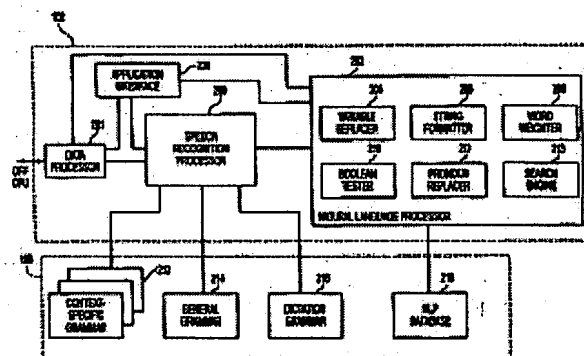
more >>

Report a data error here

Abstract not available for JP2002524806T

Abstract of corresponding document: WO0014728

A system and method for interacting with networked objects, via a computer using utterances, speech processing and natural language processing. A Data Definition File relates networked objects and a speech processor. The Data Definition File encompasses a memory structure relating the networked objects, including grammar files and a natural language processor. The speech processor searches a first grammar file for a matching phrase for the utterance, and for searching a second grammar file for the matching phrase if the matching phrase is not found in the first grammar file. The system also includes a natural language processor for searching a data base for a matching entry for the matching phrase; and an application interface for performing an action associated with the matching entry if the matching entry is found in the database. The system utilizes context-specific grammars, thereby enhancing speech recognition and natural language processing efficiency. Additionally, the system adaptively and interactively "learns" words and phrases, and their associated meanings.



Data supplied from the esp@cenet database - Worldwide

【特許請求の範囲】

【請求項1】 ネットワーク・オブジェクトと関連付けられたネットワーク・オブジェクト・テーブルをネットワークを介して転送する工程と、上記ネットワーク・オブジェクトに対する一致エントリに関して上記ネットワーク・オブジェクト・テーブルを検索する工程と、上記ネットワーク・オブジェクト・テーブル内において上記一致エントリが発見されたなら上記一致エントリに関連付けられた動作を実施する工程と、を含む、ネットワーク・オブジェクトとの音声対話の為にコンピュータを更新する方法。

【請求項2】 前記ネットワーク・オブジェクト・テーブルはウェブ・サイトから読み出される、請求項1記載の方法。

【請求項3】 前記ネットワーク・オブジェクト・テーブルは、複数のネットワーク・オブジェクトに対して各ネットワーク・オブジェクト・テーブルを記憶する箇所から読み出される、請求項1記載の方法。

【請求項4】 前記ネットワーク・オブジェクト・テーブルは、文脈固有文法も含むダイアログ定義ファイル内に含まれる、請求項1記載の方法。

【請求項5】 前記ネットワーク・オブジェクト・テーブルは、自然言語プロセッサ・データベースに対する各エントリも含むダイアログ定義ファイル内に含まれる、請求項1記載の方法。

【請求項6】 前記ネットワーク・オブジェクト・テーブルは、文脈固有文法および自然言語プロセッサ・データベースに対する各エントリも含むダイアログ定義ファイル内に含まれる、請求項1記載の方法。

【請求項7】 ネットワーク・オブジェクトと関連付けられたダイアログ定義ファイルであってネットワーク・オブジェクト・テーブルを含むダイアログ定義ファイルを転送するネットワーク・インタフェースと、

上記ネットワーク・オブジェクトに対する一致エントリに関して上記ネットワーク・オブジェクト・テーブルを検索するデータ・プロセッサと、

上記ネットワーク・オブジェクト・テーブル内において上記一致エントリが発見されたなら上記一致エントリに関連付けられた動作を実施するアプリケーション

Equivalent to Priority  
Literature

(11)特許出願公表番号  
特表2002-524806  
(P2002-524806A)

(43)公表日 平成14年8月6日(2002.8.6)

(12)公表特許公報(A)

(19)日本国特許庁(JP)

(51)Int.Cl.	識別記号	FI	予備審査請求 有	(全 49 頁)
G 0 6 F 17/28	3 4 0	G 0 6 F 17/28	予備審査請求 有	
3/16	3 4 0	3/16		
13/00	5 5 0	13/00		

審査請求 未請求 予備審査請求 有 (全 49 頁)

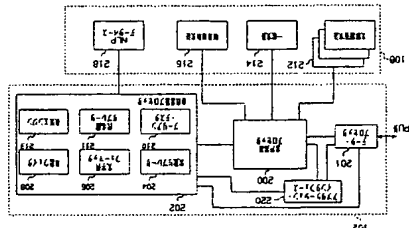
(21)出願番号	特願2000-569391(P2000-569391)	(71)出願人	ワン ボイス テクノロジーズ インコーポレーテッド
(86) (22)出願日	平成11年9月8日(1999.9.8)		アメリカ合衆国カリフォルニア州92122
(85)優先出願番号	平成13年3月9日(2001.3.9)		サン ディエゴ スイート240 グリーン
(86)出願出願番号	PCT/US99/20447		ウィッチ ドライブ 6333
(87)国際公開番号	WO00/14728 Prior art literature		ウィーバー、ディーン
(87)国際公開日	平成12年3月16日(2000.3.16)	(72)発明者	アメリカ合衆国 92131 カリフォルニア
(31)優先権主張番号	09/150, 459		州 サンディエゴ サイバラス テラス
(32)優先日	平成10年9月9日(1998.9.9)		プレイス 11422
(33)優先権主張国	米国 (US)	(74)代理人	弁理士 殿田 博宣 (外1名)
(31)優先権主張番号	09/166, 198		ファーム(参考) 58091 M15 C12 C14 C312 C003
(32)優先日	平成10年10月5日(1998.10.5)		DM3
(33)優先権主張国	米国 (US)		

最終頁に続く

(54)【発明の名称】 音声認識および自然言語処理を使用したネットワーク用対話型ユーザ・インタフェース

(57)【要約】

発願、音声処理および自然言語処理を使用してコンピュータを介してネットワーク・オブジェクトと対話するシステムおよび方法。データ定義ファイルは、ネットワーク・オブジェクトおよび音声プロセッサを関連付ける。データ定義ファイルは、文法ファイルを含むネットワーク・オブジェクトを自然言語プロセッサに対して関連付けるメモリ構造を包含する。音声プロセッサは、上記発願に対する照会対象語句に関して第1文法ファイルを検索し、もし上記第1文法ファイル内で上記照会対象語句が見えられれば上記照会対象語句に関して第2文法ファイルを検索する。上記システムはまた、上記照会対象語句に対する一致エントリに関してデータベースを検索する自然言語プロセッサ；および、もし上記一致エントリが上記データベース内で発見されたら一致エントリに関連付けられた動作を実施するアプリケーション・インタフェースも含む。上記システムは文脈固有文法を利用することにより、音声認識および自然言語処理の効率を高める。更に上記システムは、単語および語句ならびにそれらの関連意味を適合的かつ対話的に“学



ン・インタフェースと、

を含む、ネットワーク・オブジェクトとの音声対話の為にコンピュータを更新するシステム。

【請求項8】 ネットワーク・オブジェクトと関連付けられたダイアログ定義ファイルであってネットワーク・オブジェクト・テーブルを含むダイアログ定義ファイルを発見する工程と、

上記ダイアログ定義ファイルを読み出す工程と、

上記ネットワーク・オブジェクトに対する一致エントリに関して上記ネットワ

ーク・オブジェクト・テーブルを検索する工程と、

上記ネットワーク・オブジェクト・テーブル内において上記一致エントリが発見されたなら上記一致エントリに関連付けられた動作を実施する工程と、

を含む、ネットワーク・オブジェクトとの音声対話の為にコンピュータを更新する方法。

【請求項9】 前記ダイアログ定義ファイルはウェブ・サイトから読み出される、請求項8記載の方法。

【請求項10】 前記ダイアログ定義ファイルは複数のネットワーク・オブジェクトに対するダイアログ定義ファイルを記憶する箇所から読み出される、請求項8記載の方法。

【請求項11】 前記ダイアログ定義ファイルは記憶媒体から読み出される、請求項8記載の方法。

【請求項12】 前記ダイアログ定義ファイルは文脈固有文法を含む、請求項8記載の方法。

【請求項13】 前記ダイアログ定義ファイルは自然言語プロセッサ・データベースに対するエントリを含む、請求項8記載の方法。

【請求項14】 前記ダイアログ定義ファイルは文脈固有文法と自然言語プロセッサ・データベースに対するエントリとを含む、請求項8記載の方法。

## 【発明の詳細な説明】

### 【0001】

(発明の背景)

#### 1. 発明の分野

本発明はネットワーク・コンピュータ用ユーザ・インタフェースの音声認識に関する。より詳細には本発明は、音声認識および自然言語処理を使用してユーザとコンピュータが対話する新規な方法およびシステムに関する。本発明は、1998年9月10日出願されて“音声認識および自然言語処理を使用した対話型ユーザ・インタフェース”と称された米国特許出願第09/150,459号の一部継続出願である。

### 【0002】

#### 1.1. 関連技術の説明

コンピュータが普及するにつれ、コンピュータの理解およびコンピュータとの対話が多くの人々に取り困難であることが明らかとなっている。コンピュータ操作の為にユーザは、原始的なコマンドおよび直感的でない手順を学ぶべき場合が多い。例えば殆どのパーソナルコンピュータは、概してメニュー主導であるウィンドウズ（登録商標）式のオペレーティング・システムを使用する。故にユーザは、どのメニュー・コマンドもしくは一連のコマンドが所望結果を与えるのかを学ぶ必要がある。

### 【0003】

更に、コンピュータに対する習用の対話はキーボードもしくはマウスなどの手動入力デバイスの故に遅くなることも多い。多くのコンピュータ・ユーザは熟練タイピストではない。その結果、これらの手動入力デバイスによりコンピュータにコマンドおよび単語を入力するには相当の時間が掛かる。故に、コンピュータ、ならびに、ウェブ・サイトなどのネットワーク・オブジェクトと対話するの容易かつ高速で更に直感的な方法が必要なのは明らかである。

### 【0004】

コンピュータとの対話の為に提案されたひとつの方法は、音声認識である。音声認識は、人間の音声を音響的に検出すると共に検出された音声を一連の単語へ

と翻訳すべく協働するソフトウェアおよびハードウェアを必要とする。当業界において公知である如く音声認識は、ハードウェアが検出した音を、音素 (phoneme) と称される分割不能な更に小さな音へと分解することで機能する。各音素は別個の音の単位である。例えば “those” という単語は3個の音素から成り；第1の音素は “th” 音であり、第2の音素は “o” 音であり、第3の音素は “s” 音である。音声認識ソフトウェアは、検出された各音素を記憶辞書からの既知の各単語と一致させるべく試行する。音声認識システムの一例は、1998年11月8日に発行されて “音声認識装置および方法” と称され且つ Dragon Systems, Inc. に譲渡された米国特許第4,783,803号に与えられるが、該米国特許第は言及したことにより本明細書中に援用する。現在では、Dragon Systems, Inc. および International Business Machines Inc. などの会社から市販された多くの音声認識ソフトウェア・パッケージがある。

**[0005]**

これらの音声認識ソフトウェア・パッケージもしくはシステムのひとつの制限は、それらが通常はコマンドおよび制御もしくは聴音書取 (dictation) 機能のみを実施することである。故にユーザは依然として、コンピュータを操作する為にはコマンドの用語集を学ぶ必要がある。

【000】

これらの音声認識システムに対して提案された改良は、検出された単語を自然言語処理システムを用いて処理することである。自然言語処理は一般的に、検出された単語の文法的関係および相対的文脈を分析することにより各単語の概念的“意味”（例えば、話し手が意図している意味）を決定する必要がある。例えば1989年12月12日に発行されると共に“自然言語テキストに対するパーサ（PARSER）”と称され且つInternational Business Machines Inc.に譲渡された米国特許第4,887,212号は言及したことにより本明細書中に援用するが、該米国特許第は単語分離、形態学的分析、辞書検索および文法分析を使用して入力単語ストリームを解析する方法を教示している。

**[000]**

音声認識とともに自然言語処理を使用すると、話された単語を使用して、キーボードもしくはマウスなどの手動入力よりも強力なコンピュータの操作ツールが提供される。しかし乍ら従来の自然言語処理システムのひとつの欠点は、音声認識システムにより検出された単語の正しい“意味”を自然言語処理システムが決定し得ない可能性があることである。斯かる場合にユーザは通常、次の試行ではその自然言語処理システムが正しい“意味”を決定することを期待してその語句を言い換えたり言い直したりする必要がある。これは、ユーザが文もしくはコマンドの全体を言い直す必要があるときに相当の遅延に繋がるのは明らかである。従来のシステムの別の欠点は、音声認識に必要な処理時間が法外に長くなり得ることである。これは主として、処理されるべき大量の情報と比較して処理ソースの速度が有限だからである。例えば従来の多くの音声認識プログラムにおいては検索される辞書ファイルのサイズの故に、発語 (utterance) の認識に要する時間は長い。

【0008】

従来の音声認識／自然言語処理システムの更なる欠点は、それらが対話型でないことで新たな状況に対処し得ないことである。未知のもしくは新たなネットワーク・オブジェクトにコンピュータ・システムが遭遇したとき、コンピュータとオブジェクトとの間には新たなリレーションシップが形成される。故に従来の音声認識／自然言語処理システムは、それまで未知であったネットワーク・オブジェクトにより課された新たなリレーションシップに起因する状況には対処し得ない。結果として、コンピュータとの会話形式対話は不可能である。ユーザは、完全な概念をコンピュータに伝達する必要がある。ユーザはまた断片的な文を話すことが出来ない、と言うのも、（以前の発話の意味に依存する）それらの断片的な文の意味が失われるからである。

## [000]

必要なのは、上述の各欠点を回避する音声認識および自然言語処理を活用した、コンピュータとの対話型ユーザ・インタフェースである。

(發明の概要)

本発明は、発語、音声処理および自然言語処理を使用してコンピュータと対話する新規で優れたシステムおよび方法である。概略的に上記システムは音声プロセッサを含むが、該音声プロセッサは、上記発語に対する照合対象語句に関して第1文法ファイルを検索し、且つ、もし上記第1文法ファイル内で上記照合対象語句が発見されなければ上記照合対象語句に関して第2文法ファイルを検索する。上記システムはまた、上記照合対象語句に対する一致エントリに関してデータベースを検索する自然言語プロセッサ、および、もし上記一致エントリが上記データベース内で発見されたならば該一致エントリに関連付けられた動作を実施するアプリケーション・インタフェースも含む。

#### 【0010】

好適実施例において上記自然言語プロセッサは、上記一致エントリが上記データベース内で発見されなければ上記データベース、上記第1文法ファイルおよび上記第2文法ファイルの少なくともひとつを上記照合対象語句により更新する。

#### 【0011】

上記第1文法ファイルは文脈固有文法ファイルである。文脈固有文法ファイルとは、特定主題に相当に関連する単語および語句を含むファイルである。上記第2文法ファイルは一般文法ファイルである。一般文法ファイルとは、文脈に鑑みて解釈の不要な単語および語句を含むファイルである。すなわち、上記一般文法ファイル内の単語および語句は一切の起源文脈に属さない。本発明では上記一般文法ファイルを検索する前に上記文脈固有文法ファイルを検索することから、ユーザは更に会話的な形式でコンピュータと意志疎通でき、その場合に話された単語は、もし上記文脈固有文法ファイル内で発見されれば、最も最近に話された主題に鑑みて解釈される。

#### 【0012】

本発明の更なる側面において上記音声プロセッサは、もし上記照合対象語句が上記一般文法ファイル内で発見されなければ上記照合対象語句に関して聴音書取文法を検索する。上記聴音書取文法は、一般的な単語および語句の大規模な用語集である。上記文脈固有文法および一般文法を最初に検索することにより音声認識時間が相当地に短縮されることが期待されるが、これは、上記文脈固有文法およ

び一般文法は上記聴音書取文法よりも物理的に小さいからである。

#### 【0013】

本発明の別の側面において上記自然言語プロセッサは、上記データベースを検索する前に上記照合対象語句の少なくとも一個の単語を置換する。この置換は、上記照合対象語句の少なくとも一個の単語をワイルドカードで置換する上記自然言語プロセッサ内の変数リプレースにより達成され得る。上記語句内の(“単語変数”と称される)一定の単語をワイルドカードで置換することにより、上記データベース内のエントリ数は相当地に減少され得る。更に上記自然言語プロセッサ内の代名詞サブステイテュータは上記照合対象語句内の代名詞を固有名称に置換することにより、上記データベース内にユーザ特有事実が記憶されるのを許容する。

#### 【0014】

本発明の別の側面において文字列フォーマッタは上記データベースを検索する前に上記照合対象語句をテキスト・フォーマットする。また単語ウェイトは、上記データベースを検索する前に個々の単語の相対有意性に従い上記照合対象語句内の個々の単語を加重する。これらの工程によれば上記データベースの更に高速かつ正確な検索が許容される。

#### 【0015】

上記自然言語プロセッサ内の検索エンジン、上記一致エントリに対する確信値を生成する。また上記自然言語プロセッサは上記確信値をスレッシュホールドと比較する。ブーリアン・テストは、上記照合対象語句からの必要個数の単語が上記一致エントリ内に存在するか否かを決定する。このブーリアン・テストは、上記検索エンジンにより戻される結果を検証する役割を果たす。

#### 【0016】

上記照合対象語句からの必要個数の単語が上記一致エントリ内に存在しなければ、曖昧さを排除すべく上記自然言語プロセッサはユーザに対し、上記一致エントリが上記発語の正しい解釈であるか否かを問い質す。上記一致エントリが上記発語の正しい解釈でなければ、上記自然言語プロセッサはユーザに対し付加的情報を読み質す。この付加的情報により、上記データベース、上記第1文法ファイ

ルおよび上記第2文法ファイルの少なくともひとつが更新される。この様にして本発明は付加的な発語の意味を適合的に“学習”することにより、上記ユーザ・インタフェースの効率を高める。

#### 【0017】

上記音声プロセッサは、引き続き発語に対する引き続き照合対象語句に関連付けられた文脈固有文法を有効化して検索する。これにより、最も適切な単語および語句が最初に検索されて音声認識時間を短縮することが確実となる。

#### 【0018】

概略的に本発明は、ウェブ・ページなどのネットワーク・オブジェクトとの音声対話の為にコンピュータを更新する方法を含む。最初に、上記ネットワーク・オブジェクトを上記音声対話システムに関連付けるネットワーク・オブジェクト・テーブルがネットワークを介して上記コンピュータへと転送される。上記ネットワーク・オブジェクト・テーブルの場所は、特定のインターネット・ウェブ・サイトにて、または、複数のネットワーク・オブジェクトに対するネットワーク・オブジェクト・テーブルを記憶する統合箇所にて、上記ネットワーク・オブジェクト内に埋め込まれる。上記ネットワーク・オブジェクト・テーブルは、上記ネットワーク・オブジェクトに一致するエン트리に関して検索される。上記ネットワーク・オブジェクトにエントリが一致すると、テキスト音声スピーカから音声発生され、文脈固有文法ファイルが使用され、または、自然言語プロセッサ・データベースが使用されるなどの動作が行われ得る。上記ネットワーク・オブジェクト・テーブルは、ダイアログ定義ファイルの一部とされ得る。ダイアログ定義ファイルは、文脈固有文法ファイルを、または、自然言語プロセッサ・データベースに対するエントリを、または両者を含み得る。

#### 【0019】

本発明の別の側面において、ネットワーク・インタフェースは上記ネットワーク上からダイアログ定義ファイルを転送する。上記ダイアログ定義ファイルはネットワーク・オブジェクト・テーブルを含む。データ・プロセッサは、上記ネットワーク・オブジェクトに一致するテーブル・エントリに関して上記ネットワーク・オブジェクト・テーブルを検索する。この一致テーブル・エントリが発見さ

れたなら、アプリケーション・インタフェースは上記一致エントリにより特定された動作を実施する。

#### 【0020】

本発明の別の側面においては、ネットワークと関連付けられたダイアログ定義ファイルが発見されて読み出される。上記ダイアログ定義ファイルは、ウェブ・サイト、記憶媒体などの種々の場所から、または、複数のネットワーク・オブジェクトに対するダイアログ定義ファイルを記憶する場所から読み出され得る。上記ネットワーク・オブジェクトに一致するテーブル・エントリを発見すべく、上記ダイアログ定義ファイル内に含まれたネットワーク・オブジェクト・テーブルが検索される。上記一致エントリは上記ネットワーク・オブジェクトと関連付けられた動作を定義し、その動作はその後上記システムにより実施される。ネットワーク・オブジェクト・テーブルの他にも、上記ダイアログ定義ファイルは文脈固有文法を、自然言語プロセッサ・データベースに対するエントリを、または両者を含み得る。

#### 【0021】

本発明の特徴、目的および利点は添付図面に關して以下に示された詳細な説明から更に明らかとなるが、同一の参照符号は各図を通して対応している。

#### (好適実施例の詳細な説明)

次に、図1の好適なコンピュータ・システム100の機能ブロック図を参照して本発明を開示する。図1においてコンピュータ・システム100は、中央処理ユニット(CPU)102を含んでいる。CPU 102は、図3A乃至図3Dに關して本明細書中で記述される方法を実施すべく適切にプログラムされた当業界では公知の任意の汎用マイクロプロセッサもしくはマイクロコントローラとされ得る。上記CPUをプログラムするソフトウェアは記憶媒体108から、又は、コンピュータ・ネットワークを介した別の場所から読み取られ得る。例えばCPU 102は、Intel社などにより製造されたPentium IIプロセッサなどの習用のマイクロプロセッサとされ得る。

#### 【0022】

CPU 102は、ディスプレイ104、手動入力器106、記憶媒体108

、マイクロホン110、スピーカ112、データ入力ポート114およびネットワーク・インタフェース116などの複数の周辺装置と通信する。ディスプレイ104は、ユーザに対してイメージおよびテキストを視覚表示する当業界で公知のCRT、LCD画面、タッチパネルまたは他のモニタなどの視覚的ディスプレイとされ得る。手動入力器106は、習用のキーボード、キーパッド、マウス、トラックボール、または、データの手动入力のために当業界で公知の他の入力デバイスとされ得る。記憶媒体108は、磁気ディスク・ドライブ、フロッピー（登録商標）・ディスク・ドライブ、CD-ROM、シリコン・メモリ、または、データの記憶および読出しのために当業界で公知の他のメモリ・デバイスとされ得る。重要な点として、記憶媒体108はCPU 102から離間して配置され得ると共に、ローカル・エリア・ネットワーク（LAN）もしくはワイド・エリア・ネットワーク（WAN）もしくはインターネットなどのネットワークを介してCPU 102に接続され得る。マイクロホン110は、CPU 102に対してオーディオ信号を提供する当業界で公知の任意の適切なマイクロホンとされ得る。スピーカ112は、CPU 102からのオーディオ信号を再生する当業界で公知の任意の適切なスピーカとされ得る。マイクロホン110およびスピーカ112は、適切なならばデジタル/アナログおよびアナログ/デジタル変換回路を含む。データ入力ポート114は、RS-232、ユニバーサル・シリアル・バスなどのデータ・プロトコルを使用して外部付属品とインタフェースする当業界で公知の任意のデータ・ポートとされ得る。ネットワーク・インタフェース116は、コンピュータ・ネットワークを介してファイルを通信もしくは転送する当業界で公知の任意のインタフェースとされ得るが、斯かるネットワークの例としては、TCP/IP、イーサネット（登録商標）、またはトークン・リング・ネットワークが挙げられる。これに加え、一定のシステム上にてネットワーク・インタフェース116はデータ入力ポート114に接続されたモデムから成り得る。

#### 【0023】

故に図1は、コンピュータ・システム100の各機能要素を示している。コンピュータ・システム100の各要素は、上述の如き適切な既製の構成要素とされ

得る。本発明は、音声を使用してコンピュータ・システム100との人的対話を行う方法およびシステムを提供する。

#### 【0024】

図8に示された如くコンピュータ・システム100は、コンピュータ・ネットワークの集合であるインターネット700に接続され得る。インターネット700に接続する為に、コンピュータ・システム100はデータ入力ポート114に接続されたモデムであるネットワーク・インタフェース116、または、当業界で公知の他の任意の方法を使用し得る。ウェブ・サイト710、他のコンピュータ720および記憶媒体108もまた、当業界で公知の斯かる方法によりインターネットへと接続され得る。

#### 【0025】

次に図2を参照すると、図2はCPU 102および記憶媒体108の拡大機能ブロック図を示している。図2の各機能要素は全体的にソフトウェアもしくはハードウェアまたは両者にて具現され得ることは理解される。ソフトウェア実施例の場合、そのソフトウェアは記憶媒体108にて、または、コンピュータ・ネットワークを介した代替的箇所に読み取られ得る。CPU 102は、音声認識プロセッサ200、データ・プロセッサ201、自然言語プロセッサ202およびアプリケーション・インタフェース220を含む。データ・プロセッサ201は、ディスプレイ104、記憶媒体108、マイクロホン110、スピーカ112、データ入力ポート114およびネットワーク・インタフェース116とインタフェースする。データ・プロセッサ201によれば上記CPUは、これらのソースからデータを発見して読み出し得る。自然言語プロセッサ202は更に、変数リプレーサ204、文字列フォーマッタ206、単語ウェイト（word weight）208、ブーリアン・テスト210、代名詞リプレーサ211および検索エンジン213を含んでいる。記憶媒体108は、複数の文脈固有文法ファイル212、一般文法ファイル214、聴音書取文法216、および自然言語処理（NLP）データベース218とを含んでいる。好適実施例において各文法ファイル212、214および216は、ユーザが話す言語の構造を記述するバックス・ナウア記法（BNF）ファイルである。BNFファイルは言語の

構造を記述するものとして当業界で公知であることから、BNFファイルの詳細は本明細書中では論じない。BNFファイルのひとつの利点は、語句もしくは単語のシーケンスを記述する上でこれらの単語シーケンスの全ての組合せを明示的に記述することなく、階層ツリー状構造が使用され得ることである。故に、該好適実施例においてBNFファイルを使用することにより記憶媒体108内の各ファイル212、214および216の物理的サイズは最小化され、以下に記述される如くこれらのファイルが有効化されて検索され得る速度が高まる。但し代替実施例においては、他のファイル構造が使用される。

#### 【0026】

次に図3A乃至図3Dのプロローチャートを参照し、図2の各機能要素の動作および相互作用を記述する。図3Aにおいて、フローはブロック300にて開始し、音声プロセッサ200に対して発語 (utterance) を提供する。発語とは、始まりおよび終わりを有する一連の音であり、話された単語を一個以上含む得る。話された単語を捕捉するマイクrohホン110はブロック300の工程を実施し得る。代替的に、発語はデータ入力ポート114を介して記憶媒体108から音声プロセッサ200へと提供され得る。好適には、上記発語は公知の“wav”オーディオ・ファイル・フォーマットなどのデジタル・フォーマットである。

#### 【0027】

フローは判断302に進み、其処で音声プロセッサ200は、文脈固有文法212のひとつが有効化されているか否かを決定する。もし文脈固有文法212が有効化されていれば、ブロック304にては各文脈固有文法212が検索される。好適実施例において各文脈固有文法212は、起源文脈 (parent context) に関する単語および語句を含むBNFファイルである。概略的に、文脈が対象領域である。例えばパーソナルコンピュータに適用可能な本発明の実施例において、文脈の例は“ニュース (news)”、または“気象 (weather)”もしくは“株 (stocks)”などである。斯かる場合に各文脈固有文法212は、これらの文脈の夫々に対応するコマンド、制御単語、説明、修飾語もしくはパラメータを含むことになる。各文脈を使用すると、情報の形式

に対して階層構造が提供される。各文脈およびそれらの用法は、NLPデータベース218を参照して以下で更に説明する。

#### 【0028】

もし文脈固有文法212が有効化されていれば、文脈固有文法212はブロック300にて提供された上記発語に対する一致に関して検索される。但し文脈固有文法212が未だ有効化されていなければ、フローはブロック308へと進み、其処で一般文法214が有効化される。

#### 【0029】

好適実施例において一般文法214は単語および語句を含むBNFファイルであり、斯かる単語および語句自体は起源文脈に属しないが、文脈固有文法ファイル212が存在する関連文脈を有し得る。換言すると、一般文法214中の単語および語句は階層的な文脈構造のルートに存在し得る。例えばパーソナルコンピュータに適用可能な一実施例において、一般文法214はコマンドおよび制御語句を含むことになる。

#### 【0030】

ブロック310にて一般文法214は、ブロック300にて提供された発語に対する照合対象単語もしくは語句に関して検索される。ブロック312にては、一致が見られたか否かに依存して判定が為される。もし一致が見られなければ、ブロック314にて聴音書取文法216が有効化される。好適実施例において聴音書取文法216は各単語のリストを含むBNFファイルであり、各単語自体は起源文脈も関連文脈も有さない。例えばパーソナルコンピュータに適用可能な一実施例において、聴音書取文法216は一般的辞書に類似した一般的単語の比較的に大きなリストを含んでいる。

#### 【0031】

ブロック316にて上記聴音書取文法は、ブロック300にて提供された発語の各単語に対する照合対象単語に関して検索される。判定ブロック318にて照合対象単語が発見されなければ、ブロック320にてはディスプレイ104に視覚的エラー・メッセージが選択的に表示され、または、スピーカ112により可聴エラー・メッセージが選択的に再生される。而して上記プロセスは、ブロック

300にて音声プロセッサ200に対し別の発語が提供されるまで停止する。

#### 【0032】

故に上記記述から理解され得る如く、音声プロセッサ200に対して発語が提供されたとき、もし有効化された文脈固有文法ファイル212があれば、その文脈固有文法ファイル212が先ず検索される。有効化された文脈固有文法ファイル212内で一致が無ければ、一般文法214が有効化され且つ検索される。もし一般文法214でも一致しなければ、聴音書取文法316が有効化され且つ検索される。

#### 【0033】

好適実施例において音声認識プロセッサ200が文脈固有文法212または一般文法214のいずれかを検索しているとき、それは“コマンド/制御”モードに在ると称される。このモードにおいて音声認識プロセッサ200は、発語全体を全体として上記文法中の各エントリと比較する。これと対照的に、音声認識プロセッサ200が上記聴音書取文法を検索しているとき、それは“聴音書取”モードに在ると称される。このモードにおいて音声認識プロセッサ200は、聴音書取文法216内の各エントリに対し、上記発語を一度に一単語ずつ比較する。上記コマンド/制御モードにおいて発語全体に対する一致を検索することは、聴音書取モードにおいて一度に一単語ずつ検索することよりも一般的に高速であると期待される。

#### 【0034】

更に、任意の個々の文脈固有文法212は一般文法214よりもサイズが小さく(すなわち合計の単語および語句が少なく)、一般文法214自体は聴音書取文法216よりもサイズが小さいことも期待される。故に、有効化された任意の文脈固有文法212を最初に検索すれば、少なくとも部分的な理由としてファイル・サイズが小さい故に、一致があれば更に迅速に発見されると思われる。同様に、聴音書取文法216の前に一般文法214を検索することにより、一致があれば更に迅速に発見されると思われる。

#### 【0035】

また、文脈固有文法212および一般文法214の両者に対して適合的に付加

される本発明の機能に関して以下で更に説明される如く、これらの文法は最も一般的な発語を含んでいる。その故に、聴音書取文法216におけるよりも文脈固有文法212もしくは一般文法214における方が更に迅速に一致が発見されることが期待される。

#### 【0036】

最後に、以下で更に説明される如く、有効化された文脈固有文法212における単語および語句はユーザにより発語される可能性が更に高いものである、と言うのも、それらの単語および語句はユーザが最も最近に話した主題に関連する単語だからである。これにより、有効化された文脈固有文法212に従ってユーザの各単語の意味が解釈され乍ら、ユーザは断片的な文を使用して更に会話的な形式で話すことができる。

#### 【0037】

本発明は上述のシーケンスで検索を行うことから、予測される全ての単語および語句を含む単一の大きなリストで一度にひとつのエントリを検索するよりも更に効率的な検索を行い得る。

#### 【0038】

判定306に戻ると、文脈固有文法212内で一致が発見されたなら、図3Aに示された3つの可能な次工程が在る。有効化された文脈固有文法212内の各一致エントリに対し、音声認識プロセッサ200により行われる関連動作が在り得る。ブロック322はひとつの動作がアプリケーション・インタフェース220に指示を行い、該アプリケーション・インタフェース220が別体のソフトウェア・アプリケーションもしくはエントリに関して一定の動作を行い得ることを示している。例えばアプリケーション・インタフェース220は、Microsoftによる音声アプリケーション・プログラミング・インタフェース(SAPI)規格を使用して外部アプリケーションと通信し得る。上記外部アプリケーションに対しては例えば、特定のインターネット・ウェブ・サイトにアクセスすること、または、テキストを音声に変換して特定語句を話すこと、が指示され得る。他の動作は、図4のNLPデータベース218に関して以下で更に論じられる如く行われ得る。

## 【0039】

ブロック324は別の動作として、自然言語処理(NLP)データベース218におけるひとつの行(row)に直接的にアクセスすることにより、以下で更に記述される各自然言語処理工程をバイパスし得ることを示している。ブロック326は別の動作として、文脈固有文法306で発見された照合対象単語もしくは語句に対し、有効化された文脈の単語もしくは語句を前方付属(prepend)し得ることを示している。例えば有効化された文脈が“映画(movies)”であり且つ照合対象発語が“8時(8 o'clock).”であれば、単語“映画”が語句“8時”に前方付属されて“8時の映画(movies at 8 o'clock)”という語句が形成される。

## 【0040】

同様に、もし一般文法214内に一致が発見されればフローは、ブロック32に進み其処でアプリケーション・インタフェース220に上記の動作を行うことを指示し、または、ブロック324に進み其処でNLPデータベース内のひとつの行が直接的にアクセスされる。但し一般文法214において一致が発見されたとときに文脈の前方付属は生じない、と言うのも、上述の如く一般文法214内の各エントリはそれら自体が起源文脈を有さないからである。

## 【0041】

代替的に、ブロック301にては手動入力された単語が捕捉され、上記自然言語プロセッサへと入力される。

最後に図3Aを参照すると、各単語は手動入力器106を介して手動で入力され得る。この場合に音声認識は不要であるが、入力された単語の自然言語処理は依然として望ましい。故にフローは図3Bへと進む。

## 【0042】

図3Bのブロック328において、自然言語プロセッサ202は自然言語処理分析の為に上記語句をフォーマットする。このフォーマットは、文字列フォーマット206により達成されると共に、各単語間の二重空白を除去し、全ての文字を小文字(もしくは大文字)にし、短縮を広げる(例えば、“its”を“itis”に変更する)などのテキスト処理を含み得る。このフォーマット工程

の目的は、語句を解析に対して準備することである。

## 【0043】

フローはブロック330へと続き、該ブロックにては、NLPデータベース218へのアクセスに備えて変数リプレーサ204により関連ワイルドカード機能により一定の“単語変数(word-variable)”が置換される。本明細書中で使用される“単語変数”とは、量、日付、時間、通貨などを表す単語もしくは語句を指している。例えば一実施例において“8時にはどの映画が放映されるか(what movies are playing at 8 o'clock)”という語句はブロック330にて“時間にはどの映画が放映されるか(what movies are playing at time)”と変換されるが、この場合に“時間(time)”は任意の時間値を表すべく使用されるワイルドカード機能である。別の場合の一実施例においては、“IBMの株を100ドルで売る(sell IBM stock at 100 dollars)”という語句はブロック330にて“IBMの株をドルで売る(sell IBM stock at dollars)”と変換されるが、この場合に“ドル(dollars)”は任意のドル値を表すべく使用されるワイルドカード機能である。この工程は簡単なループにより達成され得るが、該ループは“ドル”もしくは“時(o'clock)”などの単語であるキー・トークンに関して語句を検索し、且つ、特定のワイルドカード機能により単語変数を置換するものである。置換が行われた語句内の箇所を追跡する為に配列が使用される。これにより、NLPデータベース218が検索された後で一定の位置にて上記語句へと元の単語変数を戻すべく再置換することが許容される。

## 【0044】

ブロック330にて単語変数を関連ワイルドカード機能により置換する目的は、NLPデータベース218内に存在すべきエントリの個数を減少することである。例えばNLPデータベース218は、8時(8 o'clock)、9時(9 o'clock)、10時(10 o'clock)などの別体のエントリではなく、“時間にはどの映画が放映されるか(what movies are playing at time)”という語句のみを含む。NLPデータ

ベース218は以下で更に記述される。

#### 【0045】

ブロック332にては、上記語句内の各代名詞が代名詞リプレサ211により適切な名称で置換される。例えば一実施例において、代名詞“私(I)”、“私の(my)”もしくは“私のもの(mine)”は話し手の氏名で置換される。この工程の目的は、ユーザがNLPデータベース218内にユーザ特有の事実を記憶してアクセスするのを許容することである。例えば“誰が私の子供ですか(who are my children)”という文は、“誰がDeanの子供ですか(who are Dean's children)”に変換される。この場合に“Dean”は話し手の固有名称である。此処でも、この工程は簡単なループにより達成され得るが、該ループは、代名詞に関して上記語句を検索し、発見された代名詞を適切な固有名称に置換する。置換が行われた語句内の箇所を追跡する為に配列が使用され得る。

#### 【0046】

ブロック334にては語句内の個々の単語が単語ウエイタ208により、該語句の全体的意味に対する“重要性”もしくは“有意性”に従い加重される。例えば一実施例においては、割当てられる3種の加重係数が在る。最低加重係数は、“a”、“an”、“the”および他の品詞などの単語に割当てられる。最高加重係数は、上記語句の意味に対して有意性を有すると思われる単語に対して与えられる。例えばそれらの単語とは、NLPデータベース218内の全ての動詞、名詞、形容詞および固有名称である。中間加重係数は、上記語句内の他の全ての単語に与えられる。この加重の目的は、NLPデータベース218に対する更に強力な検索を許容することである。

#### 【0047】

図4には、一実施例におけるNLPデータベース218の選択列の例が示されている。NLPデータベース218は複数の列400~410および複数の行412A~412Nを備えている。列400において各エントリは、上記NLPデータベースに対して“既知”の語句を表している。列402においては、列400における各エントリに対して必要な単語の個数が示される。列404において

は、列400における各エントリに対する関連文脈または下位文脈が示される。列408および410においては、列400における各エントリに対し、一個以上の関連動作が示される。尚、図4に示されたNLPデータベース218は本発明を教示すべく簡素化された例に過ぎないことを銘記されたい。他の実施例は、種々のエントリにより更に多くのもしくは更に少ない列を有し得る。

#### 【0048】

図3Bに戻ると、ブロック336にてNLPデータベース218は、該NLPデータベース218の列400内のエントリが上記語句内の各単語（またはそれらの同義語）のいずれかを含むか否かということと、これらの単語の相対重みとに基づき、上記語句に対する可能の一致に関して検索される。ブロック338にては、上記語句内における各単語の発生回数と、各単語の相対重みに基づき、可能の一致エントリの各々に対して確信値が生成される。データベースの加重単語検索は当業界で公知であり、且つ、バージニア州、アーリントン郡のDT Software, Inc. による製品“dtsearch”などの市販の検索エンジンにより実施され得る。同様に、同義語を使用した検索は、当業界で公知であり、且つ、ニュージャージー、プリンストンのプリンストン大学の認知科学研究所(Cognitive Science Laboratory)により開発された“WordNet”の如き公開ツールなどを使用して達成され得る。上記検索エンジンは、自然言語プロセッサ202の一体的部分とされ得る。

#### 【0049】

判定340にて自然言語プロセッサ202は、可能の一致エントリのいずれかが、一定の所定最小スレッショルドT以上である確信値を有するか否かを決定する。このスレッショルドTは、NLPデータベース218内の各エントリのいずれかに対して上記語句が一致したか否かに関する判定が為される為の容認可能最低確信値を表す。もしT以上の確信値を有する可能の一致エントリが無ければローはブロック342へと進み、選択のエラー・メッセージがユーザに対してディスプレイ104上で視覚的表示され、または、スピーカ112を介して音響的に再生される。一実施例においてユーザに対して表示されるエラー・メッセージがあれば、そのエラー・メッセージの形式は、最高確信NLPデータベース・エ

ントリ内で何個の“適合 (hit)” (すなわち、上記語句の何個の照合対象単語) が発見されたかに依存し得る。また、ゼロもしくは1個の適合であった場合には、2個以上の適合だった場合とは異なる形式のエラー・メッセージが生成される。

#### 【0050】

但しNLPデータベース218内において、当該エントリに対してT以上の確信値が存在する少なくとも一個のエントリが在れば、フローはブロック344へ進み其処で上記語句からは“ノイズ (noise)”単語が廃棄される。“ノイズ”単語としては、上記語句内の他の単語と比較して該語句の全体的意味に関してそれほど寄与しない単語が挙げられる。斯かる単語としては、冠詞、代名詞、接続詞、および同様の性質の単語が挙げられる。“非ノイズ (non-noise)”単語は、上記語句の全体的意味に関して相対的に寄与する単語が挙げられる。“非ノイズ”単語としては、動詞、名詞、形容詞、固有名称、および同様の性質の単語が挙げられる。

#### 【0051】

フローは図3Cに進みブロック346にては、最高確信一致エントリに関してNLPデータベース218の列402から非ノイズ単語要件が読出される。例えば、もし最高確信照合対象語句が行412A内のエントリ (例えば、“時間にはどの映画が放映されるか (what movies are playing at time)”) であれば、非ノイズ単語必要個数は3である。

#### 【0052】

判定348にては、上記語句からの非ノイズ単語必要個数が、NLPデータベース218から読出された最高確信エントリ内に実際に存在するか否かを決定するテストが行われる。このテストはブロック336において実施された適合形式検索 (relevance-style search) の精度の検証であり、上記語句と容認可能に一致しなくてもエントリは最小スレッシュホルドTよりも大きな確信値を生成する可能性が在ることは理解される。

#### 【0053】

判定348で実施される上記テストの性質は、ブーリアン・テスト210によ

り実施される論理的“AND”テストである。このテストは、上記語句 (もしくはその同義語) 内の非ノイズ単語の各々が上記最高確信エントリ内に実際に存在するか否かを決定するものである。もし上記最高確信エントリ内に実際に存在する十分な個数の必要単語が在れば、フローはブロック350へと進み、其処で自然言語プロセス202はアプリケーション・インタフェース220に対して列408もしくは410からの関連動作を行うことを指示する。付加的な動作列も存在し得ることは理解されよう。

#### 【0054】

例えば、最高確信エントリが行412A内のエントリであり、且つ、判定348のブーリアン・テストが、列400内のエントリ内の語句には実際に3個の非ノイズ単語が在ることを決定したなら、列408における関連動作 (例えば、映画ウェブ・サイトにアクセスする) が行われる。上記NLPデータベースにおける他の各エントリは他の関連動作を有する。例えば、最高確信エントリが行412Eにおけるエントリ (例えば、“いま何時 (what time is it)”) であれば関連動作は、自然言語プロセス202が (不図示の) テキスト/音声アプリケーションに対し、スピーカ112を介してユーザに現在時刻を音声発生することを指示することである。他の例として、もし最高確信エントリが行412Nのエントリ (例えば、“ニュースを見せて下さい (show me the news)”) であれば、第1の関連動作はインターネット上の所定のニュース・ウェブ・サイトにアクセスすることであり、且つ、第2の関連動作は (不図示の) イメージ表示アプリケーションに対し、上記ニュースに関連するイメージを表示することを指示することであり得る。別のもしくは付加的な動作も実施され得る。

#### 【0055】

同様に、もし最高確信エントリが判定348で決定された如き上記語句からの非ノイズ単語必要個数を含むならば、自然言語プロセス202は音声認識プロセス200に対し、列404の関連文脈に対して文脈固有文法212を有効化することを命令する。故に行412Aに対しては、“映画 (movies)”という文脈に関する文脈固有文法212が有効化される。故に、図3Aのブロック

300にて音声認識プロセッサ200に対して次の発語が提供されたとき、音声認識プロセッサ200は一般文法214を検索する前に“映画 (movies)”に対する有効な文脈固有文法212を検索する。前述の如く、適切な文脈固有文法212を有効化すると高速で好結果な音声認識の可能性が高まると共に、会話的な形式でコンピュータと対話するユーザの能力が強化される。

#### 【0056】

但し判定348にて、NLPデータベース218から読出された最高確信エンタリ内に上記語句からの非ノイズ単語必要個数が実際に存在しなければ、フローはブロック354に進み、其処でユーザはディスプレイ104もしくはスピーカ112を介し、最高確信エンタリが表されたか否かが問われる。例えばユーザが“今日のIBM株の売りはいくらか (How much is IBM stock selling for today)”と発語すると、NLPデータベース218内の最高確信エンタリは行412B内のエンタリである。この場合、適合率が高いが、必要単語（もしくはそれらの同義語）の個数は十分でない。故にユーザはブロック354にて、ユーザが“1998年8月28日のIBM株の値段はいくらか (what is the price of IBM stock on August 28, 1998)”を意味したか否かが問われる。

#### 【0057】

ユーザは、肯定的にもしくは否定的に応答し得る。ユーザが肯定的に応答したと判定356で決定されたなら、ブロック350にては最高確信エンタリに関連する動作が行われ、且つ、ブロック352にては関連する文脈固有文法212が有効化される。

#### 【0058】

但し判定356にてユーザが否定的に応答したならフローは図3Dへと進み、其処で、上記最高確信エンタリに関してNLPデータベース218の列404から関連文脈が読出され、且つ、ブロック360にてユーザは文脈用対話型ダイアログを使用して情報が問われる。例えばユーザが“XICOR株の今日の値段はいくらか (what is the price of XICOR stock

today)”と発語すると共にNLPデータベース218からの最高確信エンタリは行412B（例えば“日付のIBM株の値段はいくらか (what is the price of IBM stock on date)”）であったとすれば、ブロック354でユーザは、それがユーザの意図した処であるか否かが問われる。

#### 【0059】

もしユーザが否定的に応答したなら、“株 (stock)”という文脈がブロック358にて列404から読出され、ディスプレイ104およびスピーカ112を介してユーザには株文脈に対する文脈用対話型ダイアログが呈示される。斯かる文脈用対話型ダイアログは、XICOR株に対する名称および相場表示器記号に関してユーザに問う段階を含み得る。ユーザは必要情報を話すことで応答し得る。各可能的文脈に対しては、異なる文脈用対話型ダイアログが使用され得る。例えば“気象 (weather)”文脈用対話型ダイアログは、気象情報が所望される場所（例えば市）の名称をユーザに問う段階を含み得る。また“ニュース (news)”文脈用対話型ダイアログは、記事の種類、ニュース・ソース、ニュース・サイトに対するインターネットURL、または、他の関連情報をユーザに問う段階を含み得る。

#### 【0060】

上記文脈用対話型ダイアログが完了したなら、NLPデータベース218、一般文法214および文脈固有文法212はブロック362にて新情報を含むべく更新される。この様にして、ユーザは次回にてその情報を要求し、適切な一致が発見され、且つ、更なる情報をユーザに問うこと無く適切な動作が行われる。故に本発明は、ユーザにより発語された語句を認識すべく適格的に“学習”する。

#### 【0061】

本発明の一実施例においては、NLPデータベース218、文脈固有文法212、一般文法214および聴音書取文法216のひとつ以上が、各エンタリに関連する（不図示の）タイムスタンプ値も含み得る。一致エンタリが使用される毎に、そのエンタリに関連するタイムスタンプ値は更新される。定期的間隔にて、または、ユーザにより開始されたとき、一定の日付および時刻より以前のタイム

スタンプ値を有する各エントリは夫々のデータベース／文法から削除される。この様にして上記データベース／文法は古いもしくは期限経過エントリを“一掃”することで効率的サイズに保持され得る。これは、誤った一致の回避も促進する。

#### 【0062】

本発明の一実施例において音声認識および自然言語処理は、文脈依存音声式インタフェースを有するワールド・ワイド・ウェブ（“WWW”）もしくは“ウェブ”などのネットワーク・オブジェクトと対話すべく使用され得る。

#### 【0063】

図5は好適なダイアログ定義ファイル（DDF）500を示すが、該ファイル500は、テキストもしくはグラフィック・ファイル、または、好適実施例にてウェブ・ページなどであるインターネット・オブジェクトに対して上記音声認識および自然言語処理を関連させる為に必要な情報を表している。最も簡単な実施例においてダイアログ定義ファイル500はネットワーク・オブジェクト・テーブル510から成るが、該DDFは図5に示された如く付加的な文脈固有文法ファイル214および付加的な自然言語処理（NLP）データベース218も含み得る。DDF500の好適実施例は、ネットワーク・オブジェクト・テーブル510と、文脈固有文法ファイル214と、自然言語処理データベース218に対するエントリを含むファイルとを含む。これらの構成要素は圧縮されると共に、Lempel-Ziv圧縮などの当業界で公知の任意の方法によりDDFファイル500へと結合され得る。文脈固有文法ファイル214および自然言語処理データベース218は、上述の各段落で説明した。ネットワーク・オブジェクト・テーブル510は、リソースのアドレスをNLPデータベース218内の種々の動作、文法もしくはエントリに対して関連付けるメモリ・ツリー、チェーンもしくはテーブルなどのメモリ構造である。

#### 【0064】

ネットワーク・オブジェクト・テーブル510の好適実施例は図6に示される。図6はメモリ・テーブルを示すが、該メモリ・テーブルは次のものに対するエントリ列を含み得る：ネットワーク・オブジェクト520、テキスト→音声（T

スタンプ値を有する各エントリは夫々のデータベース／文法から削除される。この様にして上記データベース／文法は古いもしくは期限経過エントリを“一掃”することで効率的サイズに保持され得る。これは、誤った一致の回避も促進する。

#### 【0062】

本発明の一実施例において音声認識および自然言語処理は、文脈依存音声式インタフェースを有するワールド・ワイド・ウェブ（“WWW”）もしくは“ウェブ”などのネットワーク・オブジェクトと対話すべく使用され得る。

#### 【0063】

図5は好適なダイアログ定義ファイル（DDF）500を示すが、該ファイル500は、テキストもしくはグラフィック・ファイル、または、好適実施例にてウェブ・ページなどであるインターネット・オブジェクトに対して上記音声認識および自然言語処理を関連させる為に必要な情報を表している。最も簡単な実施例においてダイアログ定義ファイル500はネットワーク・オブジェクト・テーブル510から成るが、該DDFは図5に示された如く付加的な文脈固有文法ファイル214および付加的な自然言語処理（NLP）データベース218も含み得る。DDF500の好適実施例は、ネットワーク・オブジェクト・テーブル510と、文脈固有文法ファイル214と、自然言語処理データベース218に対するエントリを含むファイルとを含む。これらの構成要素は圧縮されると共に、Lempel-Ziv圧縮などの当業界で公知の任意の方法によりDDFファイル500へと結合され得る。文脈固有文法ファイル214および自然言語処理データベース218は、上述の各段落で説明した。ネットワーク・オブジェクト・テーブル510は、リソースのアドレスをNLPデータベース218内の種々の動作、文法もしくはエントリに対して関連付けるメモリ・ツリー、チェーンもしくはテーブルなどのメモリ構造である。

#### 【0064】

ネットワーク・オブジェクト・テーブル510の好適実施例は図6に示される。図6はメモリ・テーブルを示すが、該メモリ・テーブルは次のものに対するエントリ列を含み得る：ネットワーク・オブジェクト520、テキスト→音声（T

214で置換され且つ既存ファイルは無効化されることが示される。“文法追加”フラグ528は、既存の文脈固有文法ファイルと同時に新たな文脈固有文法ファイルが有効化されるべきことを示す。

#### 【0067】

最後に、残りの列のエントリは、“はい/いいえ”文法構造に関する。もし“はい/いいえ”フラグ530がマークされたなら、標準的な“はい/いいえ”文法が有効化される。標準的な“はい/いいえ”文法が有効化されたとき、コンピュータに対して肯定的なコマンドを話すと、コンピュータは“はい”なら行う“エントリ532内に示されたコマンドを実行する。同様に、コンピュータに対して否定的なコマンドを話すと、コンピュータは“いいえ”なら行う“エントリ534内に示されたコマンドを実行する。“はい”なら行う”列532および“いいえ”なら行う”列534内の各エントリは、コマンドと、NLPデータベース218内に記憶されたコマンドに対するポインタと、のいずれかとされ得る。例えば、行540Bに示された如く、“はい/いいえ”フラグがマークされる。コンピュータに対して“はい”などの肯定的回答が与えられると、“はい”なら行う”エントリ532内の対応コマンドが実行され；この特定の場合にエントリは数字“210”であり、これは上記NLPデータベース内の第210番目のコマンドへの参照である。“いいえ”という答えならコンピュータは上記NLPデータベース内の第211番目のコマンドを実行する。

#### 【0068】

図7Aに行くと、ワールド・ワイド・ウェブをブラウズするコンピュータなどのインターネットワークを行うコンピュータに対して音声および音響でコマンドを提供する方法およびシステムが示されている。図7A乃至図7Cの方法は、図3A乃至図3Dの方法と組合せて使用され得る。ブロック602にて、ワールド・ワイド・ウェブ・ブラウザに対してはウェブ・サイトURL（ネットワーク・オブジェクト）が提供される。上記ウェブ・ブラウザはインターネット上をナビゲートすべく使用されるプログラムであり、当業界で公知である。ブロック602においてブラウザに対してURLを提供する段階は、ユーザがURLを手動タイプ入力するか、または、選択されたウェブ・サイトURLに対する“リンク”

をユーザに選択させる如く可及的に容易である。この段階はまた、NLPデータベース218内の各エントリに関連付けられた動作に関して前述された発声コマンドの結果でもあり得る。URLが与えられたならコンピュータはブロック604にて、上記URL内で指定されたウェブ・サイトのインターネット・アドレスを決定し得るか否かを判定せねばならない。この決定プロセスは、当業界で公知のプロセスである。もしコンピュータが上記インターネット・アドレスを決定し得なければ、ブロック605にてはブラウザ画面内にエラー・メッセージが表示され、システムはその初期開始状態600へと戻る。もし上記インターネット・アドレスが決定されるなら、コンピュータはブロック606にてウェブ・ページに対する要求をウェブ・サイトへと送信する。

#### 【0069】

ブロック608にては、上記ウェブ・サイトがウェブ・ページを送信するか否かに依存して判定が為される。もし上記ウェブ・サイトが応答しなければ、または、ウェブ・ページの送信が不首尾となれば、ブロック605にてはブラウザ画面内にエラー・メッセージが表示され、システムはその初期開始状態600へと戻る。もし上記ウェブ・サイトがウェブ・ページを戻したなら、ブロック610ではブラウザ画面内にそのウェブ・ページが表示される。

#### 【0070】

判定ブロック612にてコンピュータ100は、上記ウェブ・サイトに対応するDDFファイル500が既に該コンピュータ100内に存在するか否かを決定する。もし上記DDFファイルが存在すれば、フローは図7Cへと進み、存在しなければフローは図7Bへと進む。

#### 【0071】

図7Bに行くと、もしDDFファイル500が存在しなければ上記コンピュータは、DDFファイル500の場所がURLとしてウェブ・ページのハイパー・テキスト・マークアップ言語（HTML）内にコード化されているか否かを吟味する。（HTMLは当業界で公知であることから、該言語の詳細は本明細書では論じない。）HTMLコード内へのDDFファイル場所のコード化は、

```
<meta DDF="http://www.conversationa
```

```
l s y s . c o m / C o n v e r s e l t . d d f " >
```

などの様に初期HTMLメタ・タグ内にDDFファイル場所を列挙するか、ブラウザによりサポートされるHTMLの変種へとスクリプト・タグを直接的に書き込むことで行われる：

```
<!--
<DDF="http://www.conversationalsys
.com/ConverseIt.ddf">
```

```
-->
```

もしDDFファイル場所の情報がウェブ・ページ内にコード化されていれば、その場所のインターネット・アドレスはブロック616にて決定され、ブロック626にてコンピュータはDDFファイル500の転送を要求する。

#### [0072]

代替的に、もしDDFファイル500の場所が上記ウェブ・ページ内にコード化されていないければ、DDFファイルの場所が記憶され得る幾つかの代替的場所がある。DDFファイルの場所は、ルート・ディレクトリ内の一定のファイル箇所などのウェブ・サイトにおいて、または、別のインターネット・サーバもしくは図1の記憶媒体108などの異なる中央箇所において、事前定義箇所内に記憶され得る。ブロック618および620はこれらの可能性に対してテストを行う。ブロック618は、DDFファイルが上記ウェブ・サイトに配置されているかを決定する。この工程にて上記コンピュータは、DDFファイル500の存在に関して照会を行うクエリを上記ウェブ・サイトに送信する。もし上記ウェブ・サイトにDDFファイル500が存在すれば、上記コンピュータはブロック626にてDDFファイル500の転送を要求する。もしDDFファイル500が上記ウェブ・サイトに配置されていないければ、上記コンピュータはブロック620にて上記ウェブ・サイトに関するDDFファイルの存在に関して中央箇所に照会する。もし上記DDFファイルが上記ウェブ・サイトに存在すれば、上記コンピュータはブロック626にて上記DDFファイルの転送を要求する。もしDDFファイル500が発見されなければ、ブロック622にては、ネットワーク・オブジェクト・テーブル510、上記ウェブ・サイトに関連するNLPデータベース

ース218、および、以前に訪れた一切のウェブ・サイトに対する文脈固有文法214、などの既存DDFファイルの既存構成要素が作動停止される。更にブロック624にて、上記ウェブ・サイトは非音声起動ウェブ・サイトとして処理されると共に、標準的文法ファイルのみが使用される。標準的文法ファイルは上記システム上に存在する文法ファイルであるが、上記ネットワーク・オブジェクトに関連する文脈固有文法ファイルと関連する一切の文法は除かれる。

#### [0073]

もしブロック626にてDDFファイル500が要求され乍らもその転送が不首尾であれば、ブロック622では存在する一切のDDFファイル500の一切の既存構成要素が作動停止され、且つ、ブロック624では上記ウェブ・サイトが非音声起動ウェブ・サイトとして処理され且つ標準的文法ファイルのみが使用される。

#### [0074]

もしブロック626にてDDFファイル500が要求されると共にその転送が好結果であれば、そのDDFファイルはブロック630にて一切の先行DDFファイルに置き換わる。ブロック632にては、ネットワーク・オブジェクト・テーブル510、文脈固有文法ファイル214およびNLPデータベース218などのDDFファイル500の一切の構成要素が抽出される。図2の機能要素を含め図3A乃至図3Dに示された方法を実現する為に必要なソフトウェアを獲得する為に、同様の技術が使用され得る。

#### [0075]

フローは図7Cへと移動する。ブロック634にてネットワーク・オブジェクト・テーブル510は上記コンピュータによりメモリ内に読取られる。ブロック636により決定された如く、もし上記ウェブ・ページURLが上記サイト・ネットワーク・オブジェクト・テーブル510内に存在すれば、斯かるウェブ・ページURLは図6に示された如く上記テーブルの行540A乃至540Eにより表される。もし上記ウェブ・ページに対応する行が存在しなければ、上記ウェブ・ページに対する非音声相互作用が存在し、処理は終了する。

#### [0076]

ブロック636で決定された如くもし上記サイト・ネットワーク・オブジェクト・テーブル510内にウェブ・ページURLが存在すれば、ブロック638にて上記コンピュータは、TTSフラグ522がマークされているか否かをチェックすることにより上記ウェブ・ページにテキスト音声524が関連付けられるか否かを決定する。もしテキスト音声524が在れば、それはブロック640にて音声発生され、フローは継続する。もし判定ブロック642にて決定された如く上記ウェブ・ページに関連付けられた文脈固有文法ファイルが在れば、それはブロック644で有効化され、次にブロック646でNLPデータベース218が有効化される。もし上記ウェブ・ページに文脈固有文法ファイルが関連付けられていなければ、ブロック646にてNLPデータベース218のみが有効化される。上記NLPデータベースが有効化646されたなら、上記システムは上述の如く図3A乃至図3Cの様挙動する。

#### 【0077】

要約すると、本発明はコンピュータのネットワーク用対話型ユーザ・インタフェースに対する方法およびシステムを提供する。ダイアログ定義ファイルを紹介してインターネット・オブジェクトに連結された文脈固有文法を使用することにより、本発明は音声認識時間を短縮すると共に、ユーザが会話的な形式でウェブ・ページなどのインターネット・オブジェクトと対話する能力を高める。更に、種々の文法およびNLPデータベースを適合更新することにより、本発明は対話効率を更に高める。

#### 【0078】

各好適実施例に関する上記記述は、当業者が本発明を実施もしくは使用し得る様に提供されたものである。当業者であればこれらの実施例の種々の改変例は容易に明らかであると共に、本明細書中で定義された包括的原理は発明能力を使用せずとも他の実施例に適用され得る。故に本発明は本明細書中に示された各実施例に限定されることは意図されず、本明細書中に開示された原理および新規特徴に一致する最大範囲に従うものとする。

#### 【図面の簡単な説明】

【図1】本発明と共に使用される好適なコンピュータ・システムの機能プロ

ック図。

【図2】図1のコンピュータ・システムにおける本発明のCPU 102および記憶媒体108の拡大機能ブロック図。

【図3A】コンピュータに対する対話型音声認識／自然言語処理を提供する方法を示すフローチャート。

【図3B】コンピュータに対する対話型音声認識／自然言語処理を提供する方法を示すフローチャート。

【図3C】コンピュータに対する対話型音声認識／自然言語処理を提供する方法のフローチャート。

【図3D】コンピュータに対する対話型音声認識／自然言語処理を提供する方法を示すフローチャート。

【図4】本発明の好適な自然言語処理(NLP)データベースの選択列を示す図。

【図5】本発明に係る好適なデータベース定義ファイル(DDF)を示す図。

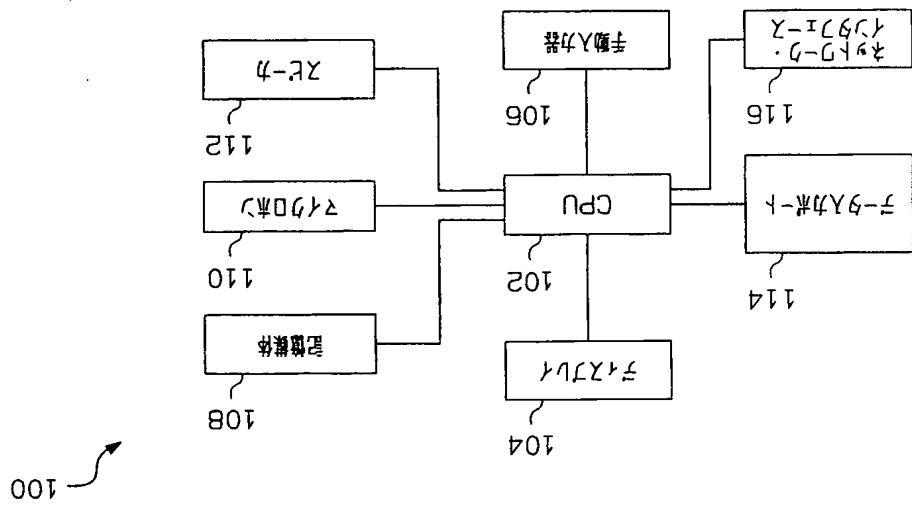
【図6】本発明の好適なネットワーク・オブジェクト・テーブルの選択列を示す図。

【図7A】ウェブ・ページなどのネットワーク・オブジェクトに対する対話型音声認識／自然言語処理のリンクを示す、本発明の方法のフローチャート。

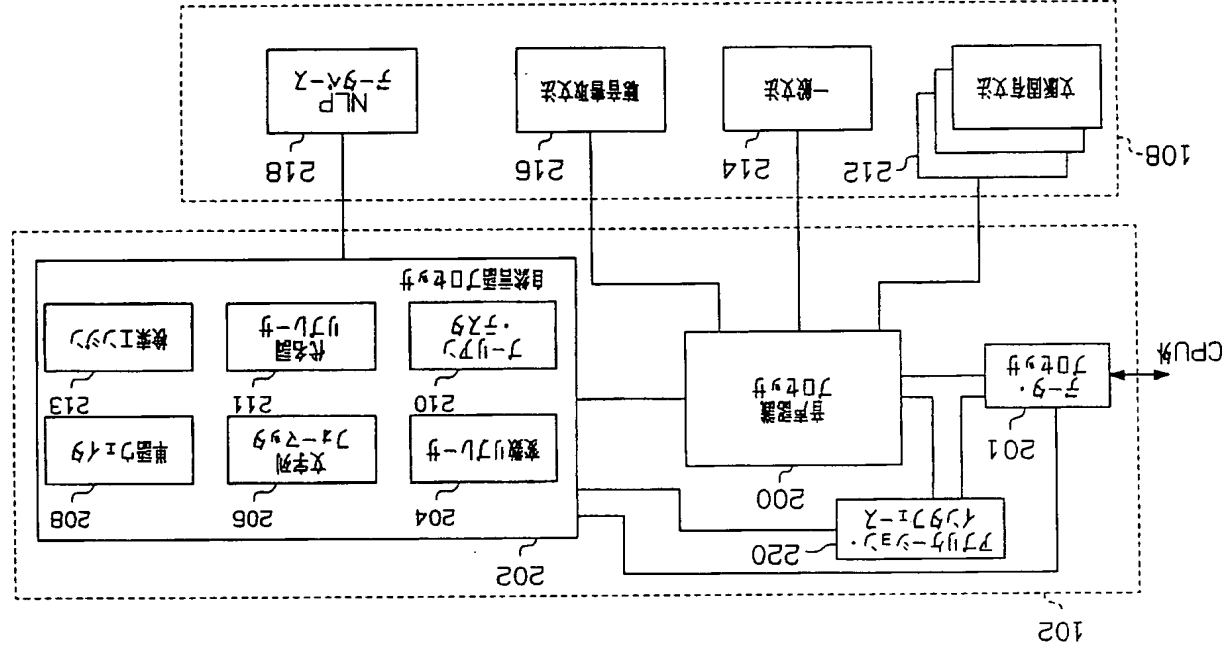
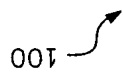
【図7B】ウェブ・ページなどのネットワーク・オブジェクトに対する対話型音声認識／自然言語処理のリンクを示す、本発明の方法のフローチャート。

【図7C】ウェブ・ページなどのネットワーク・オブジェクトに対する対話型音声認識／自然言語処理のリンクを示す、本発明の方法のフローチャート。

【図8】インターネットを介して他のコンピュータ、記憶媒体およびウェブ・サイトに接続しているコンピュータ・システムを示す図。

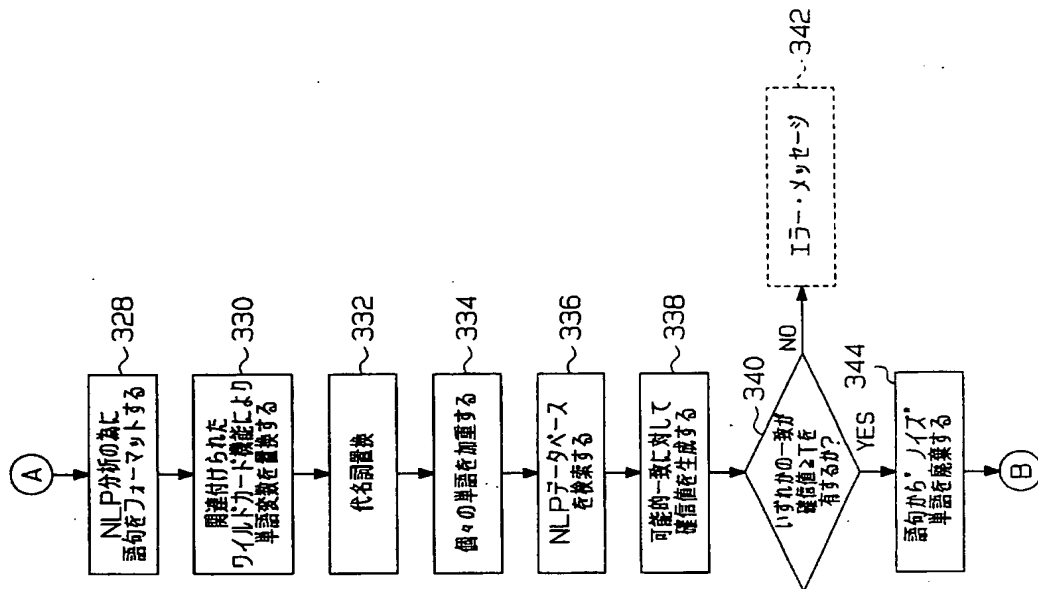


【圖 1】

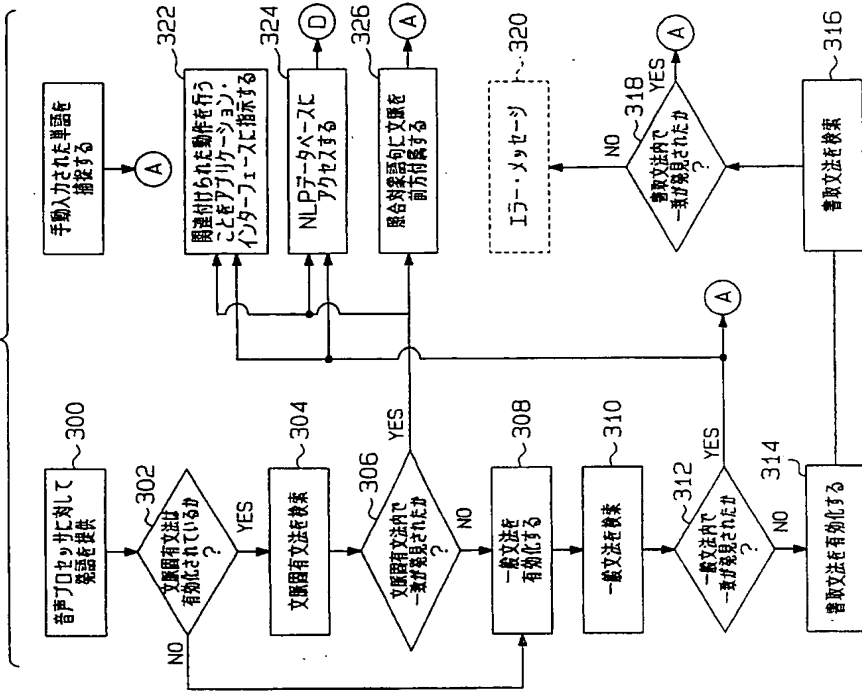


【图2】

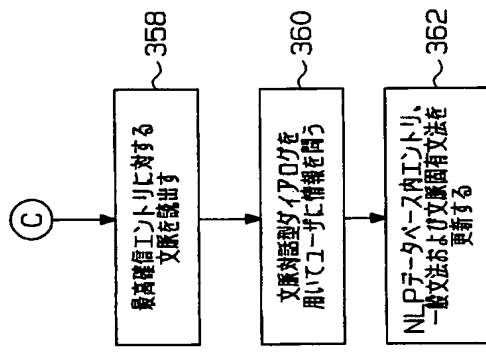
【図 3 B】



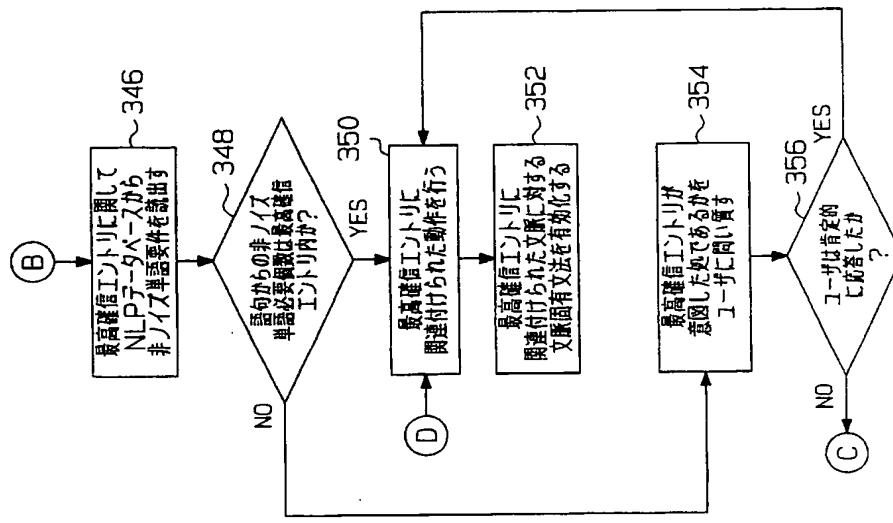
【図 3 A】



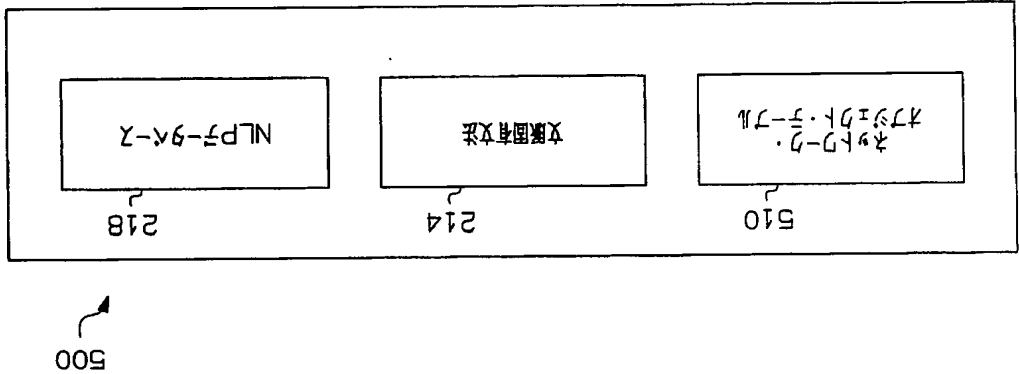
【図3D】



【図3C】



【図5】

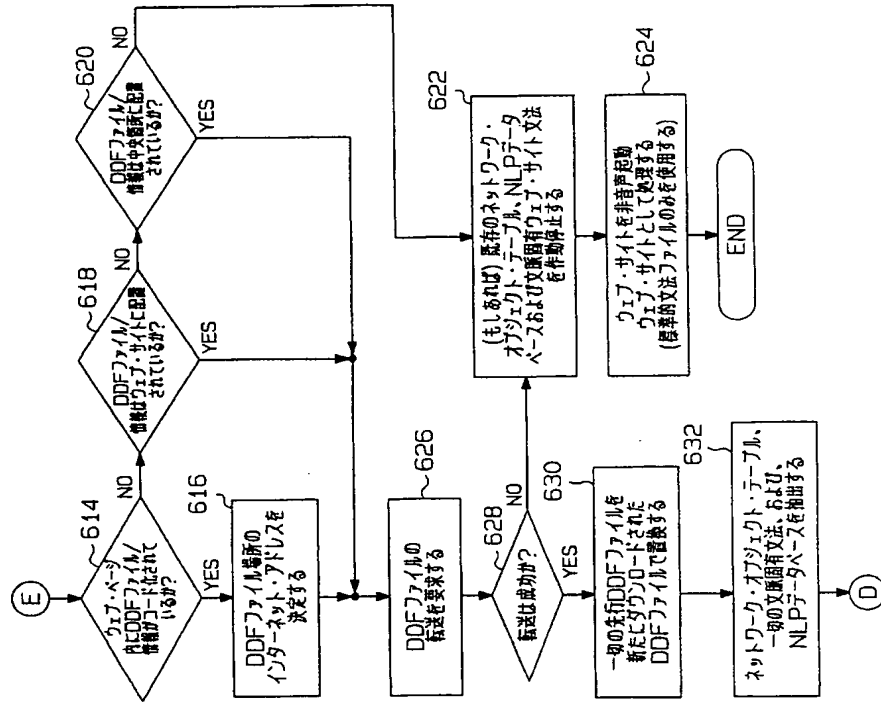


【図4】

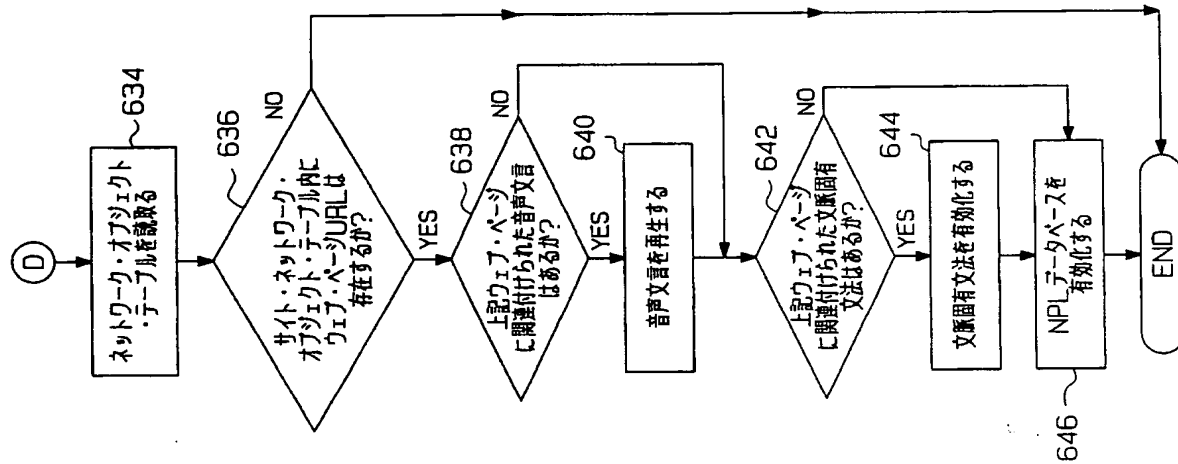
218	400	402	404	408	410
412A	時間(\$)にはどの映画が放映されるか (WHAT MOVIES ARE PLAYING AT \$TIME)	3	映画 (MOVIES)	動作1 映画データベースにアクセスする	動作2
412B	日付(\$)のIBM株の値段はいくらか (WHAT IS THE PRICE OF IBM STOCK ON \$DATA)	4	株 (STOCKS)	株価データベースにアクセスする	該当せず
412C	IBMの株を(\$)F.Lで売る (SELL IBM STOCK AT \$DOLLARS)	4	株 (STOCKS)	株価データベースにアクセスする	該当せず
412D	場所(\$)における気象はどうか (WHAT IS THE WEATHER IN \$LOCATION)	3	気象 (WEATHER)	気象データベースにアクセスする	該当せず
412E	いま何時 (WHAT TIME IS IT)	2	時刻 (TIME)	時刻データベースにアクセスする	時刻データベースにアクセスする →音声にする
412N	ニュースを見せて下さい (SHOW ME THE NEWS)	2	ニュース (NEWS)	ニュースデータベースにアクセスする	イメージを表示



【図 7 B】



【図 7 C】

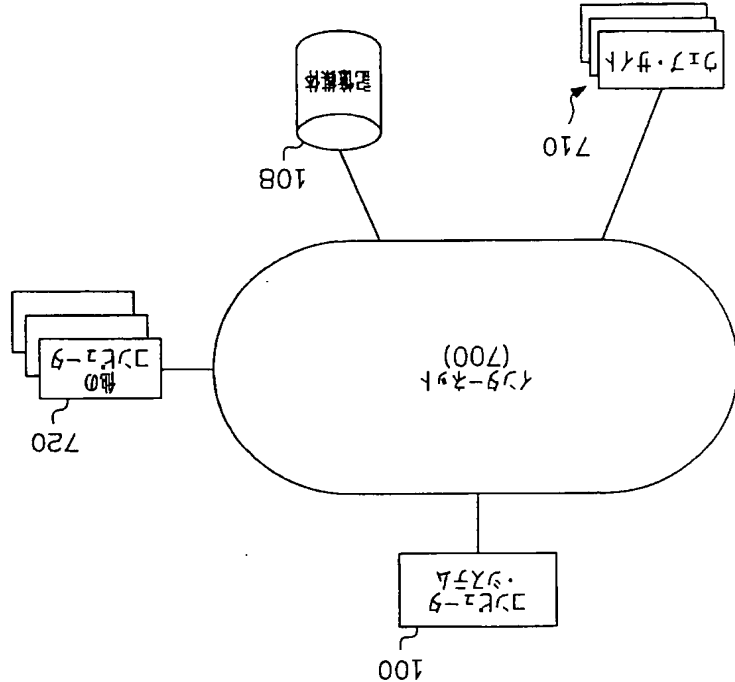


【国際調査報告】

INTERNATIONAL SEARCH REPORT

1. CLASSIFICATION OF SUBJECT MATTER IPC 7 610L 15/26 104L 29/06 606F 17/27 606F 3/16 PCT/US 99/20447	
2. FIELD OF SEARCH	
3. PRIOR ART DOCUMENTS IPC 7 610L 104L 606F	
4. DOCUMENTS CONSIDERED TO BE RELEVANT	
Category	Relevant to claim 1a
X	EP 0 854 418 A (TEXAS INSTRUMENTS INC) 22 July 1998 (1998-07-22) page 2, line 17 - line 23 page 2, line 26 - line 42 page 5, line 32 - line 52 page 6, line 20 - line 25 page 6, line 31 - line 34
Y	EP 0 747 881 A (AT & T CORP) 11 December 1996 (1996-12-11) column 4, line 12 - line 27 column 5, line 22 - line 54 column 6, line 18 - line 28
X	EP 0 747 881 A (AT & T CORP) 11 December 1996 (1996-12-11) column 4, line 12 - line 27 column 5, line 22 - line 54 column 6, line 18 - line 28
Y	EP 0 747 881 A (AT & T CORP) 11 December 1996 (1996-12-11) column 4, line 12 - line 27 column 5, line 22 - line 54 column 6, line 18 - line 28
5. FURTHER INFORMATION	
6. SUMMARY OF THE INVENTION	
7. CLAIMS	
8. ABSTRACT	
9. OTHER INFORMATION	
10. OTHER INFORMATION	
11. OTHER INFORMATION	
12. OTHER INFORMATION	
13. OTHER INFORMATION	
14. OTHER INFORMATION	
15. OTHER INFORMATION	
16. OTHER INFORMATION	
17. OTHER INFORMATION	
18. OTHER INFORMATION	
19. OTHER INFORMATION	
20. OTHER INFORMATION	
21. OTHER INFORMATION	
22. OTHER INFORMATION	
23. OTHER INFORMATION	
24. OTHER INFORMATION	
25. OTHER INFORMATION	
26. OTHER INFORMATION	
27. OTHER INFORMATION	
28. OTHER INFORMATION	
29. OTHER INFORMATION	
30. OTHER INFORMATION	
31. OTHER INFORMATION	
32. OTHER INFORMATION	
33. OTHER INFORMATION	
34. OTHER INFORMATION	
35. OTHER INFORMATION	
36. OTHER INFORMATION	
37. OTHER INFORMATION	
38. OTHER INFORMATION	
39. OTHER INFORMATION	
40. OTHER INFORMATION	
41. OTHER INFORMATION	
42. OTHER INFORMATION	
43. OTHER INFORMATION	
44. OTHER INFORMATION	
45. OTHER INFORMATION	
46. OTHER INFORMATION	
47. OTHER INFORMATION	
48. OTHER INFORMATION	
49. OTHER INFORMATION	
50. OTHER INFORMATION	
51. OTHER INFORMATION	
52. OTHER INFORMATION	
53. OTHER INFORMATION	
54. OTHER INFORMATION	
55. OTHER INFORMATION	
56. OTHER INFORMATION	
57. OTHER INFORMATION	
58. OTHER INFORMATION	
59. OTHER INFORMATION	
60. OTHER INFORMATION	
61. OTHER INFORMATION	
62. OTHER INFORMATION	
63. OTHER INFORMATION	
64. OTHER INFORMATION	
65. OTHER INFORMATION	
66. OTHER INFORMATION	
67. OTHER INFORMATION	
68. OTHER INFORMATION	
69. OTHER INFORMATION	
70. OTHER INFORMATION	
71. OTHER INFORMATION	
72. OTHER INFORMATION	
73. OTHER INFORMATION	
74. OTHER INFORMATION	
75. OTHER INFORMATION	
76. OTHER INFORMATION	
77. OTHER INFORMATION	
78. OTHER INFORMATION	
79. OTHER INFORMATION	
80. OTHER INFORMATION	
81. OTHER INFORMATION	
82. OTHER INFORMATION	
83. OTHER INFORMATION	
84. OTHER INFORMATION	
85. OTHER INFORMATION	
86. OTHER INFORMATION	
87. OTHER INFORMATION	
88. OTHER INFORMATION	
89. OTHER INFORMATION	
90. OTHER INFORMATION	
91. OTHER INFORMATION	
92. OTHER INFORMATION	
93. OTHER INFORMATION	
94. OTHER INFORMATION	
95. OTHER INFORMATION	
96. OTHER INFORMATION	
97. OTHER INFORMATION	
98. OTHER INFORMATION	
99. OTHER INFORMATION	
100. OTHER INFORMATION	

【図 8】



INTERNATIONAL SEARCH REPORT

Information on patent family members			Serial Application No. PCT/US 99/20447	
Patent document used in search report	Publication date	Patent family number(s)	Publication date	
EP 0854418 A	22-07-1998	JP 10240493 A	11-09-1998	
EP 0747881 A	11-12-1996	US 5890123 A	30-03-1999	
		CA 2177865 A	06-12-1996	
		JP 8335160 A	17-12-1996	
EP 0854417 A	22-07-1998	JP 10275162 A	13-10-1998	
DE 4440598 C	23-05-1996	NONE		
WO 9835491 A	13-08-1998	AU 5674398 A	26-08-1998	
		EP 0958092 A	24-11-1999	

From PCT/US/2002/024806 (examined) (July 1999)

INTERNATIONAL SEARCH REPORT

Serial Application No. PCT/US 99/20447	
References to state No.	
Category * (Check if document, with indication, where appropriate, of the relevant passages)	References to state No.
A EP 0 854 417 A (TEXAS INSTRUMENTS INC.) 22 July 1998 (1998-07-22) column 3, line 36 - column 4, line 2 column 5, line 1 - line 14 column 5, line 29 - line 37	1-14
A DE 44 40 598 C (SIEMENS AG) 23 May 1996 (1996-05-23) abstract column 2, line 46 - column 3, line 17	1, 2, 7-9
A WO 96 35491 A (JOHNSTON ROBERT DENIS) BRITISH TELECOM (GB) 13 August 1996 (1996-08-13) column 4, line 28 - column 5, line 8	1, 2, 7-9

From PCT/US/2002/024806 (examined) (July 1999)

## フロントページの続き

(81)指定国 EP(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG), AP(GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW

【要約の続き】

習する。